Retirement as the finishing line.

Topic modeling retirement conversations on an online forum.

Tiia Emilia Sorsa

University of Helsinki
Faculty of Social Sciences
Social and Public Policy
Master's thesis
May 2019

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto – Fakultet/Sektion – Faculty	Laitos – Institution – Department
Faculty of Social Sciences	Department of Social Sciences

Tekijä – Författare – Author
Tiia Emilia Sorsa

Työn nimi – Arbetets titel – Title
Retirement as the finishing line. Topic modeling retirement conversations on an online forum.

Oppiaine – Läroämne – Subject
Social and Public Policy

Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Master's thesis	May 2019	70 + References and Appendix 52

Tiivistelmä – Referat – Abstract

Societal norms guide the discussion on the transition to retirement. The norms dictate how and when one should retire and how one should spend retirement days. Norms are expressed in ways to speak about retirement, that is, in the narratives of retirement. The normative narratives guide individual choices and define who has retired successfully. If individual's retirement does not meet the criteria described in the narratives, they may feel they have failed.

In this Master's thesis, the retirement narratives are searched from Finland's largest online discussion forum Finland24, and the narratives are compared with the narratives found in previous studies. The data are the posts, mentioning the word retirement or pension, that were written on the forum between years 2001 and 2016. As the data consists of over 300 000 conversations threads, it is, first, thematically grouped using a machine learning method called Latent dirichlet allocation topic modeling. With the help of the model, it is possible to choose from the data only the conversations that contain relevant information on retirement narratives.

Because computational topic modeling and internet's big data have yet but few applications in the Social Science research, the second research question of the thesis is, how they can be applied in the research of this discipline. There is not yet consensus on the best practices of the method's usage, hence the analytical choices made in this thesis are described in detail. An attempt is also made to develop the interpretation of the model's results: a system is created for labelling the modelled topics and for finding the key themes among all the topics the model outputs.

As a result of the modeling, five retirement specific themes were found from the forum: Social issues, Social security system, Social development, Retirement transition, and Life and feelings. Of these themes, Retirement transition was selected for the qualitative content analysis. The discussions within the theme reveal the typically normative nature of retirement discussion on the forum. What unifies the discussions is the view that choices and chances during career define how well one succeeds in retirement. This common narrative is called in the thesis the Retirement game. Before retirement, one has to work and pay pension payments. The most widely accepted reason for retirement is achieving pension eligibility age, and those who continue working after this are seen as cheating the game, threatening the younger workers. The winner of the game is the one who survives in paid work all the way to the pension eligibility age, manages to accumulate enough pension and enjoys their freedom to control how to use their time.

By combining computational topic modeling and qualitative analysis, the thesis found retirement narratives that supplement the existing knowledge of them. The expressed norms were stronger in the discussion forum, but on the other hand, there were ways to retire that were completely against the norm. The results show that using similar data and methods it is possible to find new perspectives to existing scientific knowledge of Social Sciences' research objectives. However, topic modeling and other computational methods require interdisciplinary expertise, and further research on their best practices and application possibilities is needed.

Avainsanat – Nyckelord – Keywords
retirement transition, social norms, topic modeling, computational methods

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto – Fakultet/Sektion – Faculty	Laitos – Institution – Department
Valtiotieteellinen tiedekunta	Sosiaalitieteiden laitos

Tekijä – Författare – Author
Tiia Emilia Sorsa

Työn nimi – Arbetets titel – Title
Retirement as the finishing line. Topic modeling retirement conversations on an online forum.

Oppiaine – Läroämne – Subject
Yhteiskuntapolitiikka

Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Pro gradu -tutkielma	Toukokuu 2019	70 + lähteet ja liitteet 52

Tiivistelmä – Referat – Abstract

Yhteiskunnalliset normit ohjaavat keskustelua eläkkeelle siirtymisestä. Normit sanelevat, miten ja milloin eläkkeelle pitäisi jäädä, ja mitä eläkkeellä on sopivaa tehdä. Ne ilmenevät tavoissa puhua eläkkeestä, eli eläkenarratiiveissa. Normittuneet narratiivit ohjaavat yksilöiden valintoja ja pyrkivät määrittämään, ketkä ovat onnistuneet jäämään eläkkeelle oikein. Jos oma eläkkeelle jääminen ei vastaa onnistuneen eläkkeelle jäämisen narratiivia, yksilö voi kokea voimakasta epäonnistumisen tunnetta.

Tässä pro gradu -tutkielmassa tarkastellaan, millaisia eläkenarratiiveja Suomi24-keskustelufoorumilta löytyy ja miten ne eroavat aiemmassa tutkimuksessa löydetyistä eläkenarratiiveista. Aineistona käytetään foorumille vuosina 2001-2016 kirjoitettuja kommentteja, joissa mainitaan sana eläke. Koska aineisto käsittää yli 300 000 keskusteluketjua, se luokitellaan ensin teemoittain hyödyntäen koneoppimismenetelmää, Latent dirichlet allocation -aihemallinnusalgoritmia. Mallin avulla keskusteluista voidaan valita tarkempiin laadullisiin analyyseihin ne, jotka sisältävät olennaista tietoa eläkenarratiiveista.

Koska laskennallisen aihemallinnuksen ja suurten, itsestään muodostuvien internet-aineistojen käyttö on yhteiskuntatieteissä vähäistä, tutkielman toisena tutkimuskysymyksenä on, miten niitä voi soveltaa tällä tieteenalalla. Menetelmälle ei ole vielä muodostunut kattavia yhtenäisiä käytäntöjä, ja siksi tutkielmassa kuvataan tarkasti tehdyt analyyttiset valinnat. Tutkielmassa myös kehitetään aihemallinnuksen tulosten tulkintaa: mallinnettujen aiheiden nimeämistä ja keskeisten teemojen löytämistä kaikkien aiheiden joukosta.

Mallinnuksen tuloksena foorumilta löydettiin viisi erityisesti eläkekeskusteluun liittyvää teemaa: Yhteiskunnalliset ongelmat, Sosiaaliturvajärjestelmä, Yhteiskunnallinen kehitys, Eläkkeelle siirtyminen sekä Elämä ja tunteet. Näistä Eläkkeelle siirtymiseen sisältyviä keskusteluja analysoitiin laadullisella sisällönanalyysillä. Tulosten perusteella foorumilla käytävä eläkekeskustelu on monin paikoin normittunutta. Keskusteluja yhdistää, että uran aikana tehdyt valinnat ja sattumukset määräävät, miten hyvin eläkkeelle jäädään. Tämä narratiivi nimettiin tutkielmassa Eläkepeliksi. Eläkkeelle päästäkseen pelissä pitää tehdä työtä ja maksaa eläkemaksuja. Sopiva syy jäädä eläkkeelle on tulla lakisääteiseen eläkeikään ja ne, jotka jatkavat työntekoa tämän jälkeen, huijaavat pelissä ja ovat uhaksi nuoremmille työntekijöille. Pelin voittaa selviämällä työelämässä eläkeikään asti, kerryttämällä sopivan eläkkeen ja nauttimalla eläkkeellä saadusta oikeudesta päättää omasta ajankäytöstään.

Aihemallinnusta ja laadullista analyysiä yhdistämällä tässä tutkielmassa löydettiin aiempaa tutkimusta täydentäviä eläkenarratiiveja. Keskustelufoorumilla normit ilmaistiin toisaalta voimakkaammin kuin mitä aiemmissa tutkimuksissa on havaittu, mutta toisaalta foorumilta löytyi myös kiinnostavia, täysin normista poikkeavia toimintamalleja. Tulokset osoittavat, että vastaavia aineistoja ja menetelmiä käyttäen on mahdollista löytää uusia näkökulmia jo tutkittuihin yhteiskuntatieteellisiin tutkimuskohteisiin. Aihemallinnuksen ja muiden laskennallisten menetelmien käyttö yhteiskuntatieteissä vaatii kuitenkin poikkitieteellistä asiantuntemusta, ja jatkotutkimusta sen parhaista käytännöistä ja soveltamismahdollisuuksista tarvitaan.

Avainsanat – Nyckelord – Keywords
eläkkeelle siirtyminen, sosiaaliset normit, aihemallinnus, laskennalliset menetelmät

Table of Contents

1 Introduction

Discussing old age retirement with people, one hears all kinds of stories. Plans about retiring early and travelling, or never retiring due to pulling of job or to tight economic situation. Pensioners are said to pursue active *third ages* (see e.g. Karisto 2007) with grandchildren and new hobbies, but on the other hand, some of them are known to suffer from loneliness, boredom or alcoholism. Also the retirement system, especially the statutory retirement age, accumulation of pensions and sufficiency of pension funds, is being disputed.

The retirement stories and concerns are spread not only in one-to-one conversations but also in the news and magazines (see e.g. Koskela 2019, YLE 2017), social media (e.g. Merlino 2018) and fiction (e.g. O'Nan 2019), in self-help –literature (e.g. LaValley & Finke 2014) as well as in speeches of politicians and employees (see e.g. Taylor and Earl 2016), from which they are studied by the academics. To this academic reading of *retirement narratives,* that is, the society-wide stories of retirement that shape individuals perceptions of retirement, the current Master's thesis will add.

However, the thesis also has a second objective: to present the Social and Public Policy research with a new method of *computational topic modeling* (Blei, Ng & Jordan 2003). With this automated text analysis method, combined with qualitative content analysis, the retirement narratives are searched from the vast corpus of Finland's largest online discussion forum *Finland24*, reporting (Suomi24 Kumppanuus n.d.) to have 2,3 million monthly visitors (*Suomi24* in Finnish, author's translation).

1.1 Research questions

There are over 300 000 conversation threads from year 2001 to 2016 that mention the word *'eläke'* (*retirement* or *pension* in English) in the corpus of Finland24-forum (amount calculated by a keyword search). These conversations are used as the data of this thesis to answer **my first research questions: What issues are present when Finland24-forum users discuss retirement? What, if any, retirement narratives and norms are present in the forum?** These questions are relevant for better understanding how retirement is anticipated and experienced by individuals, which is necessary in the political design of the retirement system. The forum is its special environment where writers are protected by pseudonyms, making the analysis of the commenters' representativeness among the population impossible. Thus the thesis will not claim these would be the same retirement issues and narratives that are taking place

elsewhere in the society. Nevertheless, the forum and the comments in it reflect the thoughts of the larger society, and the concerns found from there are the concerns of real Finnish people.

The second research questions concern the mixed-methods approach that is used in the thesis: What is topic modeling and how does it work in practice? How has it been applied in Social Sciences, and how can it be used in combination with qualitative content analysis?

1.2 Key theoretical concepts and the structure of the thesis

The thesis begins by defining in Chapter 2, what are the *norms* and *narratives of retirement* in Finland. The norms, as viewed in the context of the thesis, are guidelines for individuals that define what is expected from them (Xenitidou & Edmonds 2014). The society-wide *master narratives*, that is, the culture-wide ways to speak about different issues, are where the norms have evolved, and still evolve (Lyotard 1984, Smith & Dougherty 2012). In the context of retirement, the norms and narratives told of them define, when and for what *reasons to retire* (Adams & Beehr 2003), how retirement is *expected to be* (e.g. Smith & Dougherty 2012, Kujala 2006, Frantsi 2012) and do retired people value their retirement as *successful* (e.g. Havighurst 1961, Ekerdt 2004).

The focus of this thesis is on the narratives of the *old age retirement* which is, by definition, for those that are well matured. The term is misleading however, as the current statutory retirement age is around 65 years which is not necessarily a synonym to 'old' (for further discussion of the perceptions of *old* age in Europe, see e.g. Komp, Aartsen et al. 2013). However, *old age pension* and *old age retirement* are the terms used of the type of retirement that starts when a person reaches *pension eligibility age* which is why I, too, use them in the thesis.

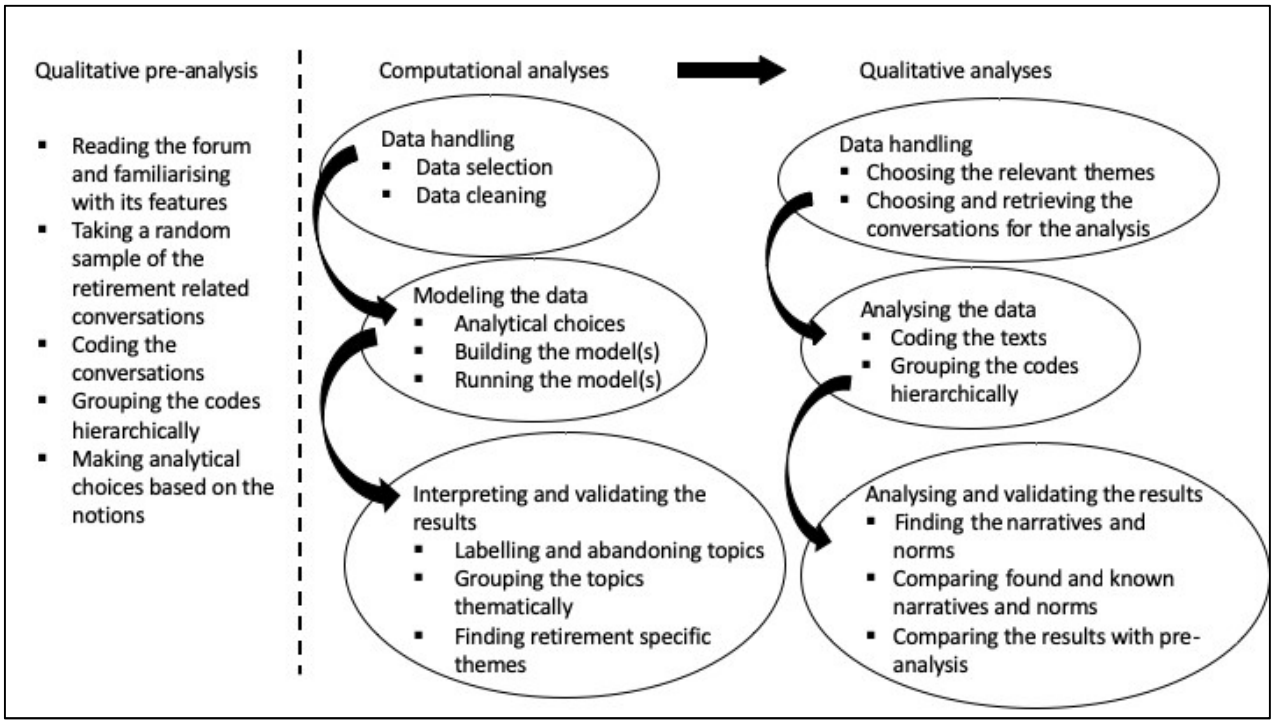
After the introduction into the norms and narratives of retirement follows the Chapter 3, a pre-analysis and comprehensive introduction of the data, i.e. Finland24-forum and the discussions in it. Why the chapter on data description spans over 12 pages is because topic modeling requires the researcher to be well familiar with the data (Grimmer & Stewart 2013).

From the Chapter 3 onwards, the words *topic* and *theme* become central. A *topic* in the way it is used in every-day speech is replaced in this thesis by the term *issue*. This is for the reason that *topic* in topic modelling has a special meaning that will be thoroughly explained in Chapter 3. For the time being, let's say a *topic* – as in topic modelling – is an output of a topic model, and that it is made of words. The "topic"

itself does not necessarily carry any meaning with it. It needs a human to interpret if the "topic", i.e. the output of the model, is really about any "issue". When it is, the topic is given a descriptive label. Together, these labelled topics may form thematic entities. Such a group of meaningful, labelled topics is in the thesis called a *theme*.

The analysis of the data is divided into the *computational* and *qualitative analyses* (Chapters 4 and 5, respectively), visualised in Figure 1:

Figure 1: The process of analyses in the thesis.



During the computational analyses, *computational topic modeling* (later abbreviated *topic modeling;* Blei, Ng & Jordan 2003) and its usage are thoroughly explained, and its "promises and pitfalls" (Grimmer & Stewart 2013) are discussed. At the time when this thesis was initialised, there were no other Social and Public Policy Master's theses using it in the University of Helsinki. Yet, topic modeling is gaining popularity in Social Sciences, and with promising results. How the method – or, strictly speaking, an application of Blei's, Ng's & Jordan's (2003) *Latent dirichlet allocation* (LDA) topic model – functions and how it was used in the thematic ordering of the forum posts will be explained on the intuitive and theoretical levels, however without going too deep into the mathematics. The results of the topic model and their interpretation answer the

second research questions, but also the issues of retirement discussion are found from the thematical grouping of the topics that ends this chapter. These results are shown and briefly discussed together with discussing the validity of the method before turning to the qualitative part of the analyses.

Topic modeling as a machine learning method may be able to categorise texts thematically into topics, however it has no capability to understand its reading. In order to find out the narratives behind the topics and answer second part of the research question, a *qualitative content analysis* is carried on based on the results of the topic modeling. The findings of the qualitative analyses indicate, combining topic modeling and qualitative analysis can produce new insights into social phenomena. The thesis concludes in the sixth chapter with discussion of future research.

As a final note on the terms used in the thesis, in computer programming, *programming* (sometimes also *coding)* means the act of writing computer programs using a programming language that a computer can compile, interpret and/or run. A *code* is a piece of a programming language. (Yatsko & Suslow 2016.) In qualitative analysis of texts, *coding* means the act of giving thematic, descriptive labels to parts of text that are under investigation. Such label is called a *code*. (Auerbach & Silverstein 2003.) In this thesis, I use the words code and coding in these both senses, and it should be evident from the context to which of these meanings the term refers.

Translations

As will be explained in Chapter Four, the output of a topic model is a list of topics of which each consists of the most probable words in it. In this thesis, the input to the model are the messages in the Finland24-forum that are in Finnish, so the output words are also in Finnish. All these words, as well as other Finnish words in the thesis, were translated to English by the author. They were translated manually, trying to find the optimal translation for every context.

In the forum, it is often impossible to say which gender the writer represents. Also, Finnish language is gender neutral in its pronouns referring to she and he, i.e. they are the same word (’hän’). Whenever gender is not specified in the message (or elsewhere in this thesis), I will use the expression ‘they’ to refer to the human being concerned.

2 Norms and narratives of retirement

To study retirement and the narratives told of it, it is necessary to define a perspective to understand them and to find them from the discussion forum. In this thesis, retirement is viewed from a perspective of it being a socially constructed system of *formal and moral norms* (Hechter & Opp 2000) that is maintained in and evolved by society-wide *metanarratives* (Lock & Strong 2010, Lyotard 1984). The norms, as seen in the thesis, are "standards of [social] acceptability" that, consciously or unconsciously, guide individuals' decision making and expectations about one self's and others' behaviour, such as the decision as of when to quit working and retire. Following a norm leads to social and self-acceptance; breaking a norm to a penalty or social disapproval. (Xenitidou & Edmonds 2014, 1.)

The formal norms of retirement are the written regulation of the retirement system, that is, the laws and contracts that individuals must obey. The moral, also called 'oughtness' norms (Hechter & Opp 2000), are the expectations of how, when and for what reasons one should retire. Oughtness norms are partially but not completely controlled by the formal regulation for the regulation does not cover all aspects of retirement. Most notably, there are no written rules about how to spend your retirement days but expectations about retirement are plenty.

The metanarratives of retirement are stories about these formal and oughtness norms of retirement, societal 'truths' that are recited from one generation to the next (Lyotard 1984). Metanarratives are one of the places where the *knowledge* of society is built, as separated from the scientific form of knowledge (Berger & Luckmann 1991). Thus, in order to know retirement, for example in the context of designing renewals of the retirement system, one must know not only the issues retirement research has proven – such as that lifting the pension eligibility age postpones actualised retirement (e.g. Takala et al. 2015) – but also the *master and personal narratives* (Smith & Dougherty 2012*)* told about norms of retirement practices. (ibid.) Presenting what is already known about different retirement narratives is the main concern of this chapter. Later in the chapters analysing the Finland24 discussions these narratives are compared to those found from the forum.

2.1 Definitions of and perspectives on retirement

There is a variety of ways to describe retirement in the retirement literature. From a holistic, societal perspective, retirement is *'an individual, organizational, and societal or*

cultural construct' (Adams & Beehr 2003, 2); *'a societal practice'* that arises from the need to *'manage succession within social groups'* (Ekerdt 2010, 69). Historically observed, retirement in its early days would mean *'a short period that began when we could not work anymore and ended when we died',* whereas today, *'retirement is a fairly long and well-deserved period in one's life when one is no longer required to pay attention to the demands of paid work'* (Adams & Beehr 2003, 3). Adam's & Beehr's approach highlights that the norms related to retirement exist, not because they are constant and inevitable, but because since the establishment of pension systems retirement has developed a status of an expected and socially accepted leisure phase in life.

There are yet other definitions that emphasize retirement as a process, such as *'[r]etirement is a life transition involving an obligatory change in how people use their time*' (Olds et al. 2018, abstract), and that it is *'a process that occurs over a span of time and that involves a series of reflections and decisions concerning timing and form of retirement...'* (Moen 2012; Solem, Syse, Furunes et al. 2016, see also Xenitidou & Edmonds 2014, 2). These 'reflections and decisions' are informed by the stories of how to retire, in other words, the retirement narratives and norms that manifest in the *macro* (the society), *meso* (the social groups) and *micro* (individuals) levels of society (Takala et al. 2015, 2. See also e.g. Komp, Aartsen et al. 2013).

The narratives are changing together with societal change, and new concepts of retirement have evolved. It has been defined as plural, constituting from the different forms of retirement such as bridge employment or part-time retirement, early retirement, voluntary retirement and perceived retirement (Adams & Beehr 2003, 1). Ekerdt (2010) also notices the less common cases of un-retirement and re-retirement, referring to the act of going back to work from retirement and retiring again.

The different definitions inform that there are multiple perspectives in approaching retirement. Although studies reviewing the retirement system are still often in the researchers' focus, the system perspective has also given way to the studies of the subjective experience of retirement process (Wang et al. 2012). Ekerdt (2010), in an opening of a special retirement issue of Journal of Gerontology, viewed the four most important retirement research objectives being "[1] *the form and timing of retirement exits,* [2] *the labor market for older workers,* [3] *the quality of pensions, and* [4] *the experience of retired life*" (ibid. 69; square brackets with numbers author's additions). Ekerdt writes from the United States' context where the retirement system and the norms linked to it are very different to their Finnish equivalents (see e.g. Adams & Beehr 2003), however the important and frequently observed themes are much alike. Looking at the articles of, for example, one of the most influential retirement research

institutes in Finland, the Finnish Centre for Pensions (Eläketurvakeskus, abbreviated ETK), Ekerdt's *'frontiers'*, as he calls them, are well covered. Among the topics ETK has studied from year 2016 to 2018 are factors that affect retirement intentions and realisations (fitting to the first of Ekerdt's (2010) frontiers'), labour market situation for older workers in different industries (the second frontier); pensioners' income (third frontier) and their subjective economic well-being (fourth frontier) (ETK 2018).

These definitions and perspectives are later needed in the analyses of the retirement discussion on the forum. As will be explained, the labelling and interpreting of the topic model's output relies on the theoretical knowledge on the subject.

2.2 Reasons for retirement

There are probably as many combinations of reasons for retirement as there are retirees. These factors that affect late career intentions to retire and the actualisation of retirement are the most covered area of retirement research (Adams & Beehr 2012, 4). Retirement is not obligatory, but its norm is so strong that almost everybody does retire. It is even so strong that the academic literature on reasons for retirement is studying early and postponed retirement, not non-retirement and retirement. Non-retirement is seen rare and extreme, suitable maybe for some entrepreneurs or other workaholics but not for "normal" people.

The norm has its origins in the pension system with its defined pension eligibility age. They "mandate (explicitly or implicitly) that older workers should leave the labour force or at least drastically reduce work hours, and that they do so at a specified age" (Adams & Beehr 2003, 9). The following sections will discuss reasons for retirement on three levels: the macro, meso and micro, which is a typical way of structuring the complex processes around the transition (see e.g. Komp, Aartsen et al. 2013, Järvensivu & von Bonsdorff 2018, 36).

Macro-level refers to the society-wide regulation of retirement, the absolute rules according to which all organisations and individuals must operate, i.e. the statutory retirement system. Socially, the most uncontested reason for retirement comes from the formal norm of retirement age. When preferred retirement age equals with *pension eligibility age*, other factors have little effect on actual retirement age (Solem, Syse, Furunes et al. 2016).

The institutions around individuals form the meso-level of retirement. Those include most importantly the families and work places whose effect on retirement decisions are manifested in the studied intentions to and reasons for retirement. Motivation to

continue at work is a good predictor of longer working lives (Oinas, Anttila and Nätti 2016, 127), and it comes from feelings of meaningfulness and enthusiasm towards work, level of independence, development possibilities, functional working environment, flexible working times, good atmosphere at work, and leadership that supports ageing at work. (Takala et al. 2015, Kyyrä & Paukkeri 2015, Kallio 2013, Tuominen & Gould 2010, Reday-Mulvay 2000.) Feel of possibly being dismissed, low support from employer and negative attitude towards work are related to increased intentions of retiring (Tuominen & Gould 2010, von Bonsdorff et al. 2010).

The macro- and meso-levels guide the retirement intentions and decisions, but it is on the micro-, that is, the individual-level where they are actualised: old age retirement is a choice (Järnefelt et al. 2017)[1]. The process of decision making is one of "hesitation and doubt" (Furunes et al. 2015, 286) that involves all the previously mentioned levels as well as health issues (e.g. Riekhoff & Järnefelt 2018, Kyyrä & Paukkeri 2015, Elovainio et al.2003), financing considerations (Takala et al. 2015) and personal preferences. The process has been tried to summarise in three phases – those of imagining retirement, considering the timing of it and realising the plans – but qualitative interviews of late career inform, the process is rather iterative. It goes back and forth the steps, interactively with other changes in life. (Ibid.)

2.3 Narratives of late career, retirement and ageing

> *Scientific knowledge does not represent the totality of knowledge; it has always existed in addition to, and in competition and conflict with, another kind of knowledge, which I will call narrative [knowledge.]*
>
> Lyotard 1984, 7.

Narratives in their broad sense are stories that are recited from one generation to the next. Knowledge of society's essential functions, "the truths", are embedded in what Lyotard (1984) and Lock & Strong (2010) call metanarratives, Smith & Dougherty (2012) master narratives, yet other writers great narratives, cultural narratives, societal

[1] However, when old age retirement is forced through, for example, disability retirement, it can hardly be described as a choice. To give the scale of the number of Finns who have retired involuntarily, there were some 214 000 Finns on disability retirement in 2016 (ETK 2017).

narratives and so on: culture-wide narratives that guide the formation and legitimacy of personal narratives. The sender (storyteller) of a narrative has a referent (source of information) from whom the sender has heard the narrative. This is the only required authority to ensure the addressee (listener) that what is being told is true and should be embraced in all actions. Addressee evaluates the narrative and, in her turn, passes it on to next listener. When a narrative spreads into social discourses, there emerges consensus within the society that legitimates the narrative. (Lyotard 1984.)

Narratives around retirement are plenty. At the time of writing this thesis, pension system sustainability is a major concern in Finnish pension discussion (see e.g. Tikanmäki et al. 2019). The differences in Finnish cohort sizes have caused pressure on increasing the pension payments or cutting the pension benefits, but is has also created a narrative that *staying at work after the pension eligibility age would be somewhat beneficial* not only to the pension system, but also for the pensioner themselves (see e.g. blogs: Lahti & Hämäläinen 2017, Piensoho 2014).

Taylor and Earl (2016) have noticed the same phenomenon affects the Australian political narratives, and it is not alien to the United States either (Achenbaum 2006). The older workers used to be spoken of as frail and better-off in retirement during the times of economic hardships. But because of labour scarcity, caused by the variation in the cohort sizes, policies aiming at prolonging careers have reshaped the narrative. The dominating *vulnerability narrative* judges those that discriminate older workers, claiming that it is the negative attitudes, not the frailty, that marginalise older people. A more recent *productivity narrative* even claims, the older workers are loyal, experienced, irreplaceable key players of productivity, and who would only lose their health if they retired. Neither of the two narratives, those of frailty and productivity, have evidence to be true, yet they are widely accepted as knowledge that spreads via the mass communication [2]. The "fictions" as the writers call the narratives, are based on factually inaccurate statements, such as that the older workers "can bring something special to your business: a lifetime of experience and knowledge that they can pass on to other workers and your customers" (Taylor and Earl 2016, 258) and that health will

[2] A good example of media reproducing the narrative of "delaying retirement supports healthy old age" was the information website of the Finnish retirement renewal of 2017 (www.elakeuudistus.fi; removed since). The pictures in there showed a red-cheeked bicyclist cycling pass a time line where there were years and ages marked. There were different amounts of apples along the line, and after her statutory retirement age, there were even more apples. My interpretation of the picture is, the bicycle, smiling bicyclist and apples advertise not only wealth but also health benefits aquired by working longer. The actual causality relation between health and retirement or continuing working remains unsolved (Taylor and Earl 2016), and economic incentives' effect on retirement decisions after pension renewal 2005 was marginal in comparison to other reasons (Takala et al. 2015).

decline faster during retirement than years at work. Needless to say, the first statement can be true or false, depending on the individual, and the second can be explained with mutual correlation of retirement and ageing to health. (Taylor and Earl 2016.)

Cultural narratives are hard to resist by individuals' personal narratives (Smith & Dougherty 2012, 474). Cultural narratives are means of power that define and maintain social norms and against which success in life is measured (Lyotard 1984). Even identity is partly constructed by relating selected cultural narratives to oneself (Kujala 2006, Frantsi 2012). In this way, cultural narratives can be harmful for those who do not fit in them. For example, Kujala (2006) recognised the social ideal of teachers' retirement to be staying *equally competent*, youthful and energetic at work until retirement. The ideal is great for those willing and capable of continuing at work to later ages, but constrains the others to feel guilty about their early retirement or downshifting plans. The old norm of frailty would function the other way round, justifying early retirement and questioning the wits of older workers. Personal narratives of the latter, in both examples, are those of failure, which neatly gives evidence that there is no absolute truth about what is a good way to retire. Objectively, early retirement means to stop working before the statutory retirement age. Subjectively, it is a whole different phenomenon depending on how one's peers react and what is the image of early retired that they see in media. According to the current Finnish oughtness norm, the retirement age also has an upper limit. In Kujala's (2006) study, there were none in the data who would have retired later than their retirement age, even though they reported willingness to it. When there are younger colleges who wait for retirement, how could you not retire as soon as you are allowed? The model story of retirement in our culture is, you are supposed to get tired with your work, fight to the statutory retirement age, and quit as soon as you can. (ibid., 157) Personal narratives of retirement are shaped by the context dependent cultural, often political, narratives.

Because cultural narratives shape personal narratives and identities, it is important that there exists a good selection of them. Post-modern era has loosened the norms and they have become more versatile (Kujala 2006, 149), at least if one has the self-esteem and view over culture to choose also from the controversial narratives. Ageing leaders', as an extreme example, could build their retirement identities from the *narratives of leadership and respectfully letting go of it*, knowing that they had the requirements for successful, joyful *third age* (see eg. Karisto & Konttinen 2004, 13) of hobbies and continued influence through networks (Frantsi 2012). People in less advantageous minorities may have few or none cultural narratives to build their retirement identities on (Van Sluytman & Torres 2014). The selection of narratives may also be that of only bad choices available. "Aged female teacher, possibly aged female

worker in any industry, faces contradictory expectations that can never be met. One has to be youthful, but not too youthful, not a grandma, but well matured, effective and competent until the very end, yet comfortably stepping aside...". The list goes on. (Kujala 2006, 151, translation from Finnish the author's). These contradictory norms form contradictory expressed identities.

Narratives of retirement shape the late career and the experience of the act of retirement, but their influence continues further yet. The time spent on retirement and the experience of ageing are also shaped by cultural narratives. The narratives of *successful retirement and ageing* (Havighurst 1961) guide the expectations of how retirement and days after it days will be like.

The norm of successful retirement varies in time and place through the narratives told of it (Achenbaum 2006). In the United States, for example, the narrative of successful retirement is mixed with that of *American dream*. Retirement is the time of "deferred gratification" (Ekerdt 2004, 4) when your own hard work has earned you the freedom to pursue your dreams. (Smith & Dougherty 2012.) Although Achenbaum (2006) claims that the end of golden era of retirement has come to an end, the relation of work and retirement is still told as "a story of working for a time — paying their dues — and then having the ability to stop working and live a life free of work in retirement" (Smith & Dougherty 2012, 461). In the modern Finnish context, expectations about retirement and pension are similar. Success in retirement means longevity, decent health and satisfaction in life (Kujala 2006) and pursuing hobbies (Frantsi 2012). Previous income, pension payments and years at work should dictate the pension benefit (Airio & Kangas 2017). On the other hand, retirement is associated with declining health and ageing (Smith & Dougherty 2012, Trethewey 2001, Rautava 2015, Gullette 1997), or the "fourth age" (Karisto 2008, 12). The "third age", in contrast, is the time, according to the narrative of successful retirement, when you have retired but are still fit and able to enjoy full life. The contrast is the fourth age, the "old old age", coincided with illnesses and tiredness.

The master narrative of *ageing as decline* is heard in the personal narratives of ageing men and women from different cultures and ethnicities (Sarvimäki 2013), and it has been part of our cultural heritage since ancient Greece (Gilleard 2007). Ageing is accepted to bring different forms of loss and isolation along with it. Career has already peaked around 50 years of age, professional networks marginalise older workers, and body escapes the standards of the youth culture. Especially for women, the double standard of ageing is burdensome: whereas in men, wrinkles mean credibility, in women, they are signs of losing sexuality and productivity. The narrative of ageing as decline is rejected by alienating oneself from the others and enterprising the self.

"Other" may grow old, but through continued learning, oneself stays track of time, and through exercise and cosmetics, the signs of ageing can be hidden and the face and body kept youthful. Together with resisting the decline-narrative, there are emancipating personal narratives of ageing as getting more experience, more mature attitude towards things in general and especially self and others, and ageing as a possibility to form new identities based on "the better self". (Kujala 2006, Trethewey 2001, Rautava 2015.)

Enterprising one's ageing is possible mainly for the privileged social groups. By separating oneself from those who really get old, the youthful older workers justify marginalising the old-looking older workers. Where there should be a movement to change the way the society and employers see the older workers, there is the acceptance and reproduction of the old norms. (Trethewey 2001.)

Havighurst (1961) suggests that success in ageing should not be measured against a social norm, but against the aged person's satisfaction with her life. If the social norm of success in retirement clashes with the person's ideal of life, and if that makes the older person feel unsatisfied, then the norm of successful ageing is wrong. For success in any societal phenomenon means that it maximises "the greatest good for the greatest number." (ibid., 8) Successful retirement and ageing are not solely in the hands of the individual, but also in those of the society that defines and maintains the norms of success.

3 Pre-analysis: Finland24 as data

> *Indeed a deep understanding of the texts is one of the key advantages of the social scientist in applying automated methods (Grimmer & Stewart 2013, 270).*

This chapter will introduce the data source used in the thesis: The Finland24 –online discussion forum. To be a chapter of data description, this one may seem excessive with its 12 pages, however the quotation below its title justifies its thoroughness. The automated method of topic modeling used in the thesis will be presented in the next chapter, and it 'indeed' requires that the researcher is well familiar with their data.

As with any method, knowing the data is a prerequisite for a successful topic modeling study. The data is supposed to hide *a thematic structure* (Blei, Ng & Jordan 2003) that the model will reveal, but before the researcher can build the model and prepare the data for it, they need to know its logical structure and have a general idea of its contents. (Grimmer & Stewart 2013, 270.) For this purpose, an introduction to the forum as data by Lagus et al. (2016) was read, the forum posts were skimmed through, keywords were used to find specific messages and a qualitative analysis of a random sample of 11 conversations was conducted. Without a comprehensive understanding, the large and messy collection of the forum posts could not have been prepared for the topic model nor could it have been analysed. Later on, the material presented in this chapter will be compared with that acquired from the topic model.

3.1 Selection of data

There are two main reasons why Finland24 forum posts were chosen as the data of the thesis. First, the initial of the study was to complement retirement research with what is called here 'non-lab data' (author's term), meaning the kind of data that did not form because researchers asked questions but because people were normally interacting without knowing they were study objects. Such self-constructing online corpora have become common study objectives in Social Sciences (Laaksonen, Matikainen, Tikka et al. 2013), yet no such studies analysing the subjective experience of retirement were found that would have utilised social media as their data.

The second reason was the availability of Finland24 posts. Unlike in interviews or surveys, the discussion in the forum is completely guided by the people who post there (except for occasional censoring by the administrators). While this denies the researcher of asking specific questions, researching such data may reveal perspectives

that would not have arisen from researcher-guided data formation and provide the study with more authentic understanding of how people, not the research community, view retirement. (Hakala & Vesa 2013, 223.) Kujala (2006) who interviewed teachers on their experiences of closing to retirement age raised the issue of how 'real' the narratives of her respondents are. The reason for this hesitation was, she did not observe the daily interaction but asked how the teachers experienced the surrounding norms, and this approach, she problematized, might have caused that some important but too obvious factors went unnoticed. Her reasoning went on, that the narratives the teachers told were those that were really meaningful to them and that made them relevant. (Ibid, 162.) But it remains open if an ethnography-like approach to hearing the narratives could reveal further information of how Finnish view retirement.

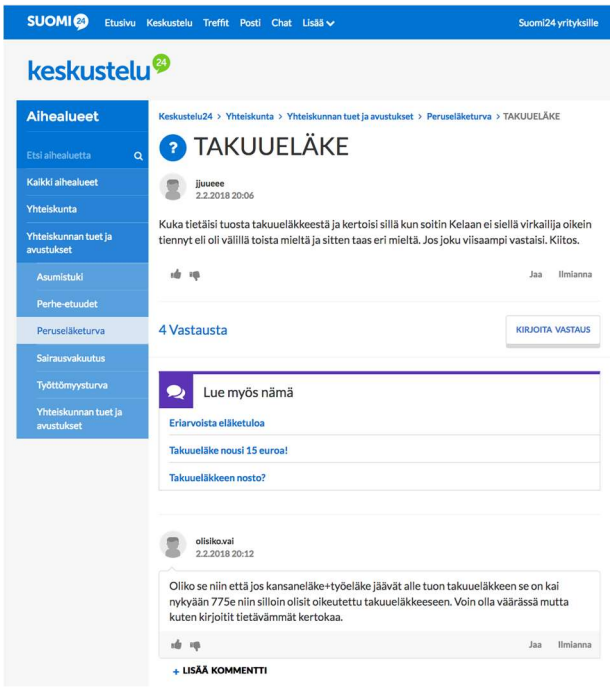
The second reason was the availability of Finland24 posts. There are vast amounts of people's online discussions in the internet, but most of them, such as Facebook-data, are owned by corporations that regulate who has access to and to which parts of their data. In 2014, a group of researchers and Aller Media Oy, the owner of Finland24, started a project called Citizen Mindscapes within University of Helsinki competition Helsinki Challenge. The following spring all conversations from year 2001 were opened for non-commercial research, today reaching to fall 2016. (Lagus, Panzar et al. 2016, Aller Media Oy 2014)

3.2 Metadata and forum structure

Finland24 is the most popular online community in Finland with 832 000 unique visitors per week. Within a month, 86 percentages of Finns using internet visit the site (Academy of Finland 2017, visiting can mean reading or posting). In spring 2016, users posted on average between 15 000 and 20 000 comments in a day (Lagus, Pantzar et al. 2016, 9). Gender distribution is near equal (women 49 percentages, men 51) and the users are of different ages: about 40 percentages are under 35 years old, 50 percentages are from 35 to 64 years and 10 percentages are over 64 years old. (Aller Media Oy 2016; data from year 2015.)

The discussion in the forum is divided into general topics such as traveling, family and economics. Under the general topics there are also sub topics, such as kids' health under family. People can start new conversation threads (also called conversations, threads, message chains and chains in the thesis) or comment existing both to general or sub topic. The user who starts the conversation gives it a heading and an entry message. Other users can comment on the entry message or on other users' comments on it.

Figure 2: An example of a conversation thread in Finland24.



The data used in the thesis are all the conversation threads in the forum from years 2001 to 2016 that mention 'retirement', 'pension' or their morphological derivative in Finnish; the exact search terms were '*eläke*', '*eläkke*', '*syytink*', '*syyting*' and '*eläköi*' (later they will be referred to these as 'retirement related words'). Unlike Ylisiurua (2017), the thesis does not limit only to the conversations that are posted on the forum under sections directly related to retirement. In her study, she noticed, the discussion can drift away from the issue of the initial writer. Thus, she proposed taking this into account when selecting the data, which is why in the present study it was chosen to use the keyword search over the whole forum instead. (ibid, 54.) Keyword search was relatively easy in this context, because the two terms – the common word 'eläke' and the old word 'syytinki' for farmers retirement arrangements – that are used of retirement nor their stems are homonyms to other meanings. The assumption that mentioning retirement would mean that the conversation was, at least partially, about retirement, was inspired also by one of the principles of the topic model used: no matter how the writers represent it, the issue they address can be read from the words they use. In a later section of the model assumptions, this logic will be described in more detail.

3.3 Conversations

Conversations are held between people who share common interests, life situations or neighbourhoods. Typically, conversation thread's structure is a question as the entry message and answers and further questions as the comments. Some users post on the forum in order to pursue their political or commercial agenda, some write just for fun. All comments are published through pseudonyms. Users can use the same pseudonym or change it as they wish. Anonymity enables people share their thoughts and feelings without a fear of stigma. On one hand, this makes people more open. On the other hand, trolls [3] and real opinions are hard to distinguish. (Lagus et al. 2016, 10.)

The 11 conversations I used to familiarize with the data were randomly selected from all conversations in the forum that mentioned a retirement related word.[4] The messages in them were then uploaded to Atlas.ti -program - a toolbox designed for qualitative analysis of texts - and coded according to their relevance to retirement or pension discussion and to things that should be taken into account in creating and analysing the model. Each document was read from 3 to 5 times, each round of reading adding more codes to it and changing the existing.

Each conversation was given from 1 to 56 codes, depending on the length and retirement-relatedness. Together, there were 111 codes in the sample corpus, which were further grouped under 10 thematic groups and one non-thematic group of codes that guided the design of the model and its analysis.[5]. All the codes were allowed to belong to one or multiple discussions as well as one or multiple code groups. Whenever there was a part in the conversation that, according to my interpretation, was not related to retirement, that part did not receive any codes.

3.3.1 Messages in the guiding sample

First conversation in the sample was under "Society"-section's subsection "Elderly care" with a heading "What is your old age like?" in the Finland24-forum. The initialising message of it raised the topics of autonomy versus care taking in the old age, euthanasia, and humane old age. The following messages discussed what is the reality behind the idealistic idea of staying at home to the very end and its alternatives, the

[3] Troll on a discussion forum means a writer or a writing that is trying to provoke disruption in the discussion.

[4] Data handling is explained in Chapter 4.

[5] By thematic I mean the codes that contained information about the writers understanding of retirement. Non-thematic means issues noted for data selection and cleaning as well as building and interpreting the model.

factors that lead to disability retirement, losing one's identity through losing their work status, helplessness and depression in old age, illness and pain, in which circumstances should there be a possibility for euthanasia and what kind of people are and should be valued in the society. The conditions in the old person's home where only visitors are the municipality care service's workers who do not have the time for proper cleaning, showering, feeding nor chatting are described in detail. On the other hand, one writer notes that there are many independent elderly people, and that it is not the responsibility of the society but that of the family to provide the elderly with decent, meaningful old age. A writer with a nickname "grandma hedgehog" notes, that the family and friends, not the formal care givers, can make the old person feel loved. The discussion's thematic codes (53 in total) include retirement as having to retire, meaning of life, in the good old times, pain, death, doctors and politics. The codes guiding the model and analysis were retiree as the writer, sharing something heard from a retiree and direct quote.

Not all the conversations were such pearls as the first one. The second thread (in Relationships: Man's logic) consisted clearly of trolls, where one of the writers presents herself as being a "79 years old mama-*pensioner*" and thus the conversation ended up into the data. The conversation was given only two codes: troll and pensioner as synonym to old person. The eighth conversation (in Society: Politics) mocked Tarja Halonen (former president of the republic of Finland) being "a leftist-pensioner-bitch". The comment was given code retiree as a mocking word, and rest of the conversation was coded as not relevant to retirement/pension. Tenth conversation (in Vehicles and traffic: Machines and heavy transportation) was full of hate, writers argue about one occupation being stupid and the workers possessing that occupation being idiots as well. One writer says that their retirement age is approaching and thus they have to degrade into lesser jobs. Despite the nasty writings, this comment is very relevant and was coded with retirement age (statutory), ageing workers and career (as the thematic codes) and mockery and retiree as the writer (as codes guiding the modeling). The last conversation (in Sports and exercise: Martial arts) went about which martial art is best. Tai chi was mocked as being park exercise for the retirees. The conversation received the codes of hobbies, age and perception of retirees as exercising in the park but was mainly coded as not related to retirement and mockery.

The third discussion (in Society: Where the world is going to) was strongly political. It discussed how the Finnish officials have regulated the MOT tests to be more frequent in Finland than what EU would allow, only so that the MOT test centres, denationalised by the National Coalition party, could flourish. Pension ceilings are mentioned, sarcastically, among other issues that are not needed to consider in Finland. The

conversation was given two thematic codes, politics and pension ceilings, and two codes guiding the modeling and interpretation, sarcasm and not related to retirement.

The fourth discussion was titled “Pensioners to work?” in Work and studies: Job announcements. It invited forum users to visit and comment on her blog. Unfortunately, the blog is not part of the data. The writer’s thoughts on how some of the elderly would like to be at work but are not asked and their experience is not appreciated, are only commented by one other writer who composes a poem on what retired life could be. Aging workers, work during retirement and widowing were among the 13 thematic codes, whereas poetry and discussion’s shift out of Finland24 were the non-thematic codes.

The fifth conversation (in Society: Politics) was a list of news from three months related to drinking and the problems is causes. Major part of the news was coded with not relevant -tag. Retired man is mentioned in two of the news as a drunken driver causing traffic accident, and in one piece, the head of the Social Insurance Institution of Finland (Kansan*eläke*laitos in Finnish) is mentioned. The thematic codes the conversation received were pensioner as a synonym to old person, substance abuse and pension institutions. Retirement is mentioned every now and then in a long conversation, urls (i.e. website addresses) and not related to retirement/pension were the non-thematic codes.

The sixth one (in Health: Drugs and addictions) was started by a desperate person who had lost 50,000 euros in gambling. The commentators share their stories of different addictions and give advice how to survive with the loans. One writer tells about a woman, retiree, who had been gambling on slot machines. The conversation was given 5 thematic codes, among them addiction in gaming and death, and sharing something heard from a retiree and not related to retirement/pension as the non-thematic codes.

Seventh conversation (in Illness: Allergies and asthma) discusses about illnesses, especially asthma, and how they complicate life. Disability retirement is seen, depending on the writer, as a mercy or a curse. Among the 17 thematic codes were retirement due to illness, pension decision, actualised retirement age, and retirement as getting and retirement as having to retire. Although the beginning of the conversation was related to difficulties in *getting* work with a severe illness and was not related to retirement (thus giving the conversation the only non-thematic code), major part of the conversation discussed about *staying* at work due to the illness.

Ninth conversation (in Family: Upbringing: General in upbringing) was highly relevant to retirement discussion. It reflected the trade-off between mothers who stay at home and raise their children properly as in good old times, and modern mothers who can’t

do so or it will mean losing career-development, income and pension. In brief, the writers agree that our system is ill-designed: accumulating pension is a prerequisite for good living in old age in today's society, due to which everyone must always work, due to which parents are not there at home for their children, due to which children lacking the necessary information take to bad habits, due to which the children get depressed, due to which the adults of tomorrow will be depressed, due to which the society will fall into decay. Throughout the conversations, pension is used as the argument why it is not a good idea to stay home, and why there should be a system of salary for mothers. Although at the time of the conversation, late November 2005, there had been a reform of the law that allowed pension accumulation during parental leave (Työeläke.fi), for the writers, this is not enough. They crave for a party that would pursue mothers' appreciation in the society. There were 42 codes in the conversation, including retirement as a central theme in the society, generation contract, values and mothers at home. Only non-thematic code was that retirement is mentioned every now and then in a long conversation.

3.3.2 Codes and themes in the sample

Although this was a relatively small sample of all the retirement related conversations, it shows in general what the conversations are like and what topics at least it is possible to find from it. Later in the analyses on the suitability of the topic model as a data selection method, the codes and themes found here will be compared to those found by the model.

The ten thematic code groups were the following (codes marked *emphasised*):

Things to do on retirement (15 codes)

8 documents mentioned things people do on retirement. Except for only *being on retirement* and *passing time*, these were *hobbies* of *television* and *sports*, valuing *active retirement* in the forms of *working during retirement*, *sex* and *intimacy* good and bad habits of *retirees as consumers* and addictions of *gambling* and *drugs* and philosophical thinking about *what does it feel like to be on retirement*. Is it like you were *always on a holiday* or just *staying at home until the very end*?

Pension system (10 codes)

Pension system related codes were present in 6 documents. Overall, *retirement system was quite central in building the society*. *The system's renewal*, *old age dependency ratio*, *sustainability gap* and *actualised* as well as *statutory retirement age* were

commented. Values mentioned were *generation contract*, time during which pensions should *accumulate* and *pensions'* size, especially *pension ceilings*.

Act of retirement (14 codes)
The act of retirement was mentioned in four documents. The reasons why people retire that were mentioned were due to *advancing disability* or *illness* reported in *rehabilitation examination*, *actualised* or *statutory retirement age* and *termination of working contract* at the final years of career that caused so called *pension pipe*. *Retirement decision* was seen both as *getting* and *having to*. Pension types mentioned were *pension due to accident* and *disability retirement*. Life *before and after retirement* was described in the contexts of illness and activeness.

Elderly care (7 codes)
One document discussed the theme of how and where the "old old age" (see also Karisto 2008) should be spent. *Old people's homes*, *enhanced services in institutions* and *home care* were problematized because *the institutions are inaccessible*, and *the role of family versus state* in providing *care for the elderly* and *humane old age* should be clearer.

Death (6 codes)
Retirement was in three conversations connected to *old age and fraility*, which in turn was related to *deceasing*. When describing the horrible conditions at home when the aim is to *stay at home until the very end*, the possibility of *euthanasia* was asked for. *Widowhood* and *meaning of life* were also addressed.

Perceptions on retirees (7 codes)
Perceptions on retirees could be found from 5 documents. Major perception was that *retiree is used as a synonym to old person* and even as *a despising nick name*. Retirees were also thought to be *armed thieves*, *lonely*, *violent* and *ill*. Also the *role of retirees* was included in the theme.

Politics and values (22 codes)
Politics and *values* was the second most code-intensive theme. Present in six documents, it covered the issues of *social security* and *welfare state* with their values of *generation contract, Nordic versus Southern European welfare model, public versus private, active ageing, feminism, value of home chores, humane old age, human value* and *well-being*. Political issues in the theme, besides those mentioned as values, were

ageing workers, euthanasia, workforce, war, voting, good old times, taxes, political organisation and *Finland versus other countries.*

Roles during life course (22 codes)
Roles during the *life course* that were mentioned in the sample were not limited to *being on retirement*. Their formation guided by *the generation contract* together with other things, these were *children, disability, family, motherhood, mother at home, parenthood, working life, work addict, working during retirement, retirees as consumers, before and after retirement, grand-parenthood, age, old age* and *widowing*. Other role-related codes were *planning the future* and *the feelings of usefulness* and *nothingness*.

Illness (8 codes)
Different *illnesses* and symptoms were the major topic in one document and three other conversations also mentioned them. They were seen as either *reasons for retirement* or things that make general life worse. Although there was a writer who was *recovering*, for many writers in the sample the *advancing disability* lead to *disability retirement*. Other codes were *doctors, pain* and *disability*.

Work and livelihood (other than retirement, 11 codes)
Work and *livelihood*, without the codes in retirement and pension system, included 11 codes. For one writer, *recession in the 90's* meant end of her *career*. *Money can buy* was mentioned as a complement to *the social security*. Some writers dreamt of *returning to work*, some noted that *home chores as work should be valued* higher. *Ageing workers* are still part of *the workforce*, and for some *work addicts*, *working life* is the life.

The ten themes show, by searching for conversations that mention retirement or pension it is possible to find both relevant and non-relevant conversations about retirement. Many of the themes are common to the current Finnish retirement research and also to the Ekerdt's (2010) classification of the frontiers of retirement research. The theme *Things to do on retirement* answers to the question of what is the style of living for retirees (Ekerdt 2010). Pension system corresponds to pension, and older workers are included in the themes of Retiring and Work and livelihood. Retirement timing is the subtlest of the four frontiers, and does not follow in any one category. However, it was addressed in seven themes, the exceptions being Death, Elderly care and Perceptions on retirees.

3.3.3 Notions on the sample

Exploring the data systematically did not only provide the study with a clue of the topics there are hidden in the conversations. It also guided the data handling as well as choice and tuning of the model, and was used as material for analysing the benefits of topic modelling in data selection.

Reading the conversations in the sample was surprising in many ways. The quality of the discussion in the first conversation raised hopes of the same suitability from other writers and look for themes recognised in the ageing research. The second reminded to look for trolls, and the third learned to only analyse the parts in the conversation that actually discussed retirement. As Levy and Franklin (2014) nicely formulated it:

> *"…public comments are rich data because commenters express all manner of concerns in unstructured ways. Comments run the gamut from technical specifications to personal stories and from thoughtful reflection to threats and name-calling. … Thus, hand-coding such documents can be a particularly difficult task; topic modeling appeals because it can uncover hidden patterns in even a varied set of documents. " (184.)*

With topic modeling it would be possible to sort these messages into analysable themes, but only if they were first controlled for the issues raised in reading the sample.

Relevant and irrelevant topics

Because of my previous studies on ageing research, and especially my Bachelor's thesis on intergenerational justice (Sorsa 2014), I found codes such as generation contract (Rawls 1971), ageing workers and Nordic versus Southern European welfare model (see e.g. Esping-Andersen 1990). Another human coder coming from a different academic context would probably have coded the same sample in a very different way. Topic modeling does not have these biases (see e.g. Levy and Franklin 2014, 184) – nor does it have the knowledge of what is relevant (see e.g. Purhonen & Toikka 2016). When retirement or pension are mentioned, it does not, in many cases, mean that the conversation, post nor even the sentence surrounding the word is related to retirement or pension. The conversation could in general be about which martial art is the best, and retirees could be mentioned only to despise one art as in the 11th conversation. Message could be a troll, as was the case in conversation number two, or the word could be part of a name of an institution as happened in the fifth conversation in the sample. Human reading the conversations is able to exclude irrelevant parts from the

analysis. Computational topic model will blindly value each co-appearing word group the same and find many topics that are not related to retirement (see also e.g. Ylä-Anttila 2018, 10). Looking at the sample, there could be more such general topics than there are specifically retirement related topics in the results of the modeling.

In order to tackle the problem that non-retirement related conversation would thus end up in the data in large quantities, the choice was made that of each conversation, only the messages were kept that mentioned a retirement related word (see Chapter 4).

Duplicates and quotes

In the conversation number ten, one of the messages or parts of it were quoted in other messages. Human reader who notices this will automatically skip the duplicates as they give no new information. The model that is told to look if words frequently co-appear, will instead see the duplicates as clear clues that the words in them belong to the same topic (Blei, Ng, Jordan 2003). The obvious direct quotes and duplicate posts were removed from the data before modeling (see Chapter 4).

Number of topics

In a sample of 11 conversations there were 110 codes that divided into 10 themes. It is not reasonable to assume that this would be all the codes and themes within the notably larger data from which the sample was taken, which means there could be thousands of topics to be found by the model. Such a number would be hard to analyse. Within the literature on topic modeling read for this thesis, the largest number of topics was 200 (Ylä-Anttila 2018). On the other hand, it would be a pity to oversimplify the data by reducing the number of topics. The question of how many topics is the right amount will be further discussed in Chapter 4.

Conversation continues elsewhere

In the conversation number four the actual discussion on the topic was held in the writer's blog that is not part of the data. Finland-24 is used as an announcement board to attract people elsewhere. The model does not follow the link, but the human reader can so do. While it is well possible to build a data retrieval tool that follows links between websites, this application was considered requiring too much effort in relation to the expected returns in the context of this study. As there were no other such conversations in the sample nor did such come up while browsing the forum, it was assumed they were rare in the forum. No such conversations ended up among those used in the qualitative analyses.

Length of the conversations varies

The longer the conversation and the more often it mentioned retirement, the more codes it was allocated with. Even the shortest conversations had more than one code and could be part of multiple themes. If the topic model was such that it classified each document into only one category, it would miss major part of the information there is to retrieve from the data (see also Levy and Franklin 2014). Also, if the model missed the different lengths of the conversations, short conversations with little information would have the same weight as longer, more informative conversations. The Latent dirichlet allocation –topic model avoids both these problems (Blei, Ng, Jordan 2003) which was a major factor in choosing the correct model.

3.4 Usage rights and ethical issues

Access and licence

Finland24 conversations are available via Kielipankki's Korp-API for anyone to "copy and redistribute the material in any medium or format" and "remix, transform, and build upon the material". However, only non-commercial use is allowed and the licensor must be credited. (The Suomi 24 Corpus n.d.) Aller Media has also given Kielipankki some of its corpora for downloading, of which the complete version of the messages posted between 1.1.2001 and 24.9.2016 was used in this thesis. These corpora are available to researchers.

Writers behind pseudonyms

The metadata there are in the corpora does not contain any personal information about the writers who use altering pseudonyms. In most cases, there were no concerns the identity of the writer could be found from the forum. However, the writer who invited Finland24 users to their website was an exception, and in order to blur their identity, the address for the website was not given in the thesis.

Although there were not such cases in the sample, it is technically possible to identify a person from the writings by combining the information they share. Knowing solely the age and workplace can already be enough. Especially if a writer always uses the same pseudonym and writes often, by combining the pieces of information in the messages they might be recognised by an acquaintance. This was not as much a problem in this thesis as it is in general, for anyone in the forum can search for the messages by pseudonyms (keskustelu.suomi24.fi, feature available on 19.12.2018). The opinions

writers share are often sensitive, and if the cover of a pseudonym is lost, they might, for example, be forced to respond to the issues in their workplace. In the thesis, all pseudonyms were removed from the messages.

A direct quote of a message can be fed into internet's search engine, and it will find its original context in the forum (own experiments with the data). Even if the message was anonymised in the research, in this way, anyone could reconnect the quote with a pseudonym and other information given in the original text. An advantage of writing in English was, all the quotations in the thesis were translations from the original Finnish posts. Translating them back, it is improbable they would be exactly the same. While this is not a plausible encryption method, it will make it at least somewhat more difficult to make the reconnection between the quote as it is in the thesis and its original source in the forum.

It is stated in the Finland24's terms of use, Aller Media has the right to use the writings of the writers as they wish (https://www.suomi24.fi/opastus/kayttoehdot). While this covers the possibility for giving them to research purposes, the writers may not know they are being studied.

4 Computational analyses

4.1 Computational topic modeling

Computational topic modeling (Blei, Ng & Jordan 2003) is machine learning method that mimics *qualitative content analysis* (see e.g. Eskola & Suoranta 1998; later abbreviated as simply *content analysis*). In both content analysis and computational topic modeling, the aim is to find a hidden thematic structure within a set of documents, *the corpus*. In content analysis, the researcher reads the documents and draws topics out of them according to their coding system. In computational analysis, the researcher describes rules for the computer, and the computer reads and classifies the documents into topics. The similarities go further: The research question and the data dictate which manual coding system or computer algorithm suits, and the researcher must know the data and the theory well in order to avoid false interpretations of the results. Both can view the topics as emerging from the corpus (called inductive analysis in content analysis and unsupervised learning in machine learning analysis) or use theory-based topics and compare the corpus with them (respectively, deductive or supervised analysis; see an example of supervised topic modeling: Blei & McAuliffe 2007). (see e.g. Mohr & Bogdanov 2013, Baumer et al. 2017, Lucas et al. 2015.)

The processing and understanding of the documents is nevertheless very different, due to which not every manual content analysis can be replaced by computational nor vice versa. Humans read deeply and classify the themes according to the meaning of the text. The process is slow, but the results - especially when the researcher is experienced and careful - are accurate and relevant (Grimmer & Stewart 2013, 268-271). Computer algorithms can be applied to huge data sets of different documents: except for textual data, topic modeling has been used for example on genetic data, images, and social networks (Blei 2012, 77). Yet, algorithms cannot understand their reading; a classic example is categorizing “time flies” into time- and insect-categories when they actually both belong to the category of time (Grimmer & Stewart 2013, 268-271). The data needs to be modified greatly before topic modeling can work on it, and the word order and grammar are dropped from the analysis. Irony, spelling mistakes, proverbs and other context-dependent meanings are thus out of their reach as well. The resulting topics that the model outputs are sensitive to even minor changes in data or model parameters (Chuang et al. 2014; note there exist other computer-assisted automated text analysis methods better suited for the above described problems, see eg. Pang & Lee 2008.) Thus, humans still outperform computers in finding themes from documents (Röder, Both & Hinneburg, 2015).

However, computational tools such as topic modeling can complement qualitative analyses of texts by showing different aspects of the data, and offering tools to summarise them (Baumer et al. 2017). The latter feature is becoming more and more valuable as the amount and the size of data available for research increases, mainly due to the popularity of publishing in social media (Grimmer & Stewart 2013, 268), as well as governments' open data projects (see e.g. Levy & Franklin 2013), computational topic modeling can help Social Sciences in acquiring more accurate and holistic understanding of such data and the phenomena in the data. Instead of subsampling the data, all of it can be taken into analysis with computational topic modeling. Research questions that were previously difficult if not impossible to answer because of the limitations in data and methods can be addressed. (See eg. Quinn et al. 2010.) The costs of the research have been analysed by Quinn et al. (ibid.), and topic modeling is notably less expensive than human reading or human coding-based analysis systems (see also Mohr & Bogdanov 2013, 561-564). With thorough validation of the model and its results, topic modeling can provide the researcher with valuable insights and summarization of the data (Chang et al. 2009, 1) and expand their view into new areas of interest (Baumer et al. 2017).

Topic modeling as applied in Social Sciences

To say that computer assisted content analysis is "new" (see Introduction) is a little inaccurate. The earliest publication I could find that describes a method analysing documents' conceptual, not the word-for-word matches, was published already in 1979 (Koll 1979). The method I use in the thesis is approximately 30 years more modern topic model called Latent dirichlet allocation, typically abbreviated LDA[6] (Blei, Ng, Jordan 2002).

This brief summarisation of articles investigates how topic modeling is applied in Social Sciences. Unlike qualitative content analysis, the method is not yet well-established among Social Scientists, and thus there is a great variance in how it is applied and problematized. There is not yet a text book where from one could study the essentials. Thus, in order to accomplish topic modeling analysis, besides studying the theoretical articles on the details of the model (such as Blei, Ng & Jordan 2003, Griffiths & Steywers 2013 and Nelimarkka 2019), it was necessary to read how other scholars have used it.

[6] In machine learning and other methodological descriptions, the abbreviation 'LDA' may refer also to Linear discriminant analysis, another statistical method. In the thesis, 'LDA' refers always to Latent dirichlet allocation.

Topic modeling has already been used as a method in many studies within Social Sciences, and it would be a subject of another study to present them all. The publications cited here are but a glance on them, presented to illustrate some of the features of how and on which subjects topic modeling is applied. The studies here were selected for their relevance within this thesis' context: either the method or its usage was well explained, or its application was close to my approach or especially clever, or it relevantly problematized some aspect of using the method. The articles are presented in Table 1:

Table 1: Studies applying topic modeling into Social Sciences.

Author	Year	Research objectives	Focus	Number of topics	Further analyses with the results of the modeling?	Type
Ahonen	2015	What effects do Big Data methods have on traditional research methods? Does budgetary legislation depend on the legislative history of the country?	Method	3	The topics' distributions between countries are briefly analysed.	article
Purhonen & Toikka	2016	What do digitalisation, big data and new computational methods offer to sociology? What are their limitations?	Explaining topic modeling	7	The topics are but briefly introduced, and further research with another model is suggested.	article
Levy & Franklin	2014	How do different actors in trucking industry frame monitoring?	Substance	8	Analyse how the different topics i.e. frames are used by different commenter groups.	article
Baumer et al.	2017	How do qualitative and computational methods differ in analysing the same data? What are the experiences of people returning to social media?	Method	10	Compares the two methods.	article
Mohr & Bogdanov	2013	What is topic modeling, what has been researched this far with them and why do they matter?	Explaining topic modeling	25	Analyses, why the topics have been formed as such	essay
Ylä-Anttila, Eranto & Kukkonen	2018	How can qualitative research benefit from the use of topic modeling? How do different actors in India and USA frame climate change?	Method	30	Short analysis of the secondary research question.	article
Ylisiurua	2017	What does automated text analysis offer to studying online data interpretatively? What are the health narratives of Finland24-forum?	Substance and method	50	Groups the topics into themes. Analyses briefly the topics and narratives found by reading the conversations within topics.	article
Nuortimo	2016	What is the discursive landscape of an online video game?	Substance and method	100	Groups the topics into three and takes exemplary topics of each group, asking why the topic was formed. Also makes a timeline analysis of the occurrences of the topics.	Master's thesis
Ylä-Anttila	2018	How counter-knowledge is expressed in online discussion?	Substance	200	Used four topics that concerned a wider theme of "knowledge" and analysed the discussion within documents of these topics	article
Brauer & Dymitrow	2014	How do the actual contents of the legislation compare to the policy's developmental goals?	Substance	300	The topics are seen as intermediate results. They were thematically grouped into three so-called axes. The relative occurrences of topics within axes' was analysed.	article
Niemi	2016	How do organisations view a project directed to erase tax avoidance?	Substance and method	25, 25, 10	Models also the topical distributions within one single consultations and within one commenter.	Master's thesis

Special effort was used to find especially the relevant Finnish publications, because modeling text means modeling language. It was hypothesised, the Finnish language would pose specific questions, and most of the Finnish publications found modelled Finnish corpora. However, word processing tools for also Finnish language have been developed, and by utilising them also the Finnish language was successfully modelled in all the studies (see next section for further information of the tools).

The articles of topic modeling - in general, and in the Table 1 - can be divided in three according to their research objectives. They are either concentrating on explaining and/or developing the model in the context of Social Sciences (see e.g. Purhonen & Toikka 2016, Mohr & Bogdanov 2013), or they approach a social phenomenon through this new method (e.g. Levy & Franklin 2014), or both (e.g. Niemi 2016). Furthermore, the latter group consists of papers whose analysis concentrate on the (labelled) output of the topic model (such as Nuortimo 2016), and those that extend topic modeling with a qualitative analysis of texts in given topics (such as Ylä-Anttila 2018). The current Master's thesis applies topic modeling similarly to this latter group. First, topic modeling is used to find the relevant issues (topics, grouped into themes) of the retirement discussion within the corpus. Then, the discussion within these themes is analysed qualitatively. The details of this process are described in the following sections.

4.2 Topic modeling Finland24-forum

4.2.1 Data handling: From raw data to vectors

The *Finland24* discussions were downloaded from the Kielipankki, resulting in 32 gigabytes of *Finland24* discussions. The corpus contained not only the writings of the people at the forum but also meta-data about, for example, the time of the writing, the writers' pseudonyms and the title of the section in which the message was published in the forum. All that was needed for the study at hand was the textual contents of the conversations that mentioned a retirement related word (see Chapter 3).

Sorting and cleaning of the data was accomplished with a programming script in Python-language (all the codes used in the thesis can be found from https://github.com/tirri/thesis). The script, first, tracked all the conversations where a retirement-related word was mentioned, either in the original entry message or in the comments that followed. If such message was found, the message was stored as a unit of analysis.

It remains unanswered within the scholars using topic modeling how to decide the correct length of a unit of analysis in general (see eg. Brauer and Dymitrow 2014, 35). It was considered to use the whole message chain in which the retirement-related word was mentioned as a unit of analysis. This approach was abandoned, however, after thorough reading of the conversation threads described in the third chapter. It is typical in the forum, that a comment or even the entry message discusses about completely different issue than the rest of the thread. Thus, if the whole message chain rather than single message was the unit of analysis, the data would have held more non-retirement related discussion. Also, the topics gathered from the data would have been unnecessary messy and their interpretation artificial. (see also Ylisiurua 2017.)

Having said that, in case more than one message within the message chain mentioned a retirement-related word, all those messages were stored together as a unit of analysis. This enabled a wider context for the retirement discussion while still excluding the messages that did not mention retirement.

As described in the third chapter, it is typical in the forum that writers quote directly on other writers. Some writers post the same message multiple times, either by accident or because they think it is relevant in different contexts. In case a direct quote of full message or a double message was in a message chain, only the original message was kept. Despite this, there were quotes where there were some parts of the original message edited, in which case the program did not recognise it as a quote. Neither did it exclude the equal messages if they were posted under different conversation trees. These faults showed in some of the topics that the model output, as it has recognised them as their own topics. Such topics were not taken into the analysis.

When the script had cut all the retirement-related messages from the raw data, the messages (or message chains that formed a unit of analysis) were stored each in a separate, human readable file into a folder. The two steps were needed to keep the conversations interpretable for the coming analysis. Unlike Ylisiurua (2017) who had to search for the specific messages from the forum itself, this indexing enabled the messages to be retrieved later. After saving the 135 992 files thus created, the transformation of the conversations continued.

The transformation, or the second step of data cleaning, prepared the data for lemmatization, that is, shortening the words to their basic form (see e.g. Plisson et al. 2004). The metadata and double messages were removed as were capitals, punctuation, website addresses and other links, numbers and special characters. If there were multiple messages in the unit of analysis, the messages were no longer separated from each other. These lists of lowercase words that the messages had become were also stored, each in a separate file, into another folder. The

corresponding original conversations and thus created bag-of-words version of it were given the same index-number for later use.

Why this bag-of-words version of a text is supposed to carry meaning is explained at the end of this section.

For lemmatization, the local copy of the latter folder was transmitted to taito.csc.fi – server where there were a lemmatization tool (a module called *finnish-process*; since removed from the available modules) for Finnish language available. For the topic model, the grammar that were required in the message's context, are irrelevant. Only the phenomenon that the word refers to is relevant. For example, *retiree* and *retirees* are different words and they are not interchangeable in a natural sentence, but they refer to the same thing: a certain group of people in certain stage of their careers. Lemmatization recognizes this and cuts the 's' from the latter. The process reduces the number of different words in the data and improves both the model efficiency and accuracy. (Plisson et al. 2004.) Lemmatization is especially relevant in a language like Finnish where suffixes and compound words are typical, and the same word in its different forms would confuse the model (Nelimarkka 2019; see, however, Purhonen & Toikka 2016, 16-18).

Figure 3: An example of a message after lemmatisation. The experssion of the writer is blurred by the process, yet it can be seen that they describe a boat called Eve that had to 'retire' because it was not maintained properly. The words have been transformed into their basic form, yet a dialectic form "mie" of the noun "I" has remained. Some compound words have been recognised and their parts have been separated by "#", but many, like "puuosa" (wooden part) have not. Lemmatization requires a complete dictionary to work, and the development of the one used in this thesis.

```
mie soudella lapsuus ajaa eve alumiinivene se olla hyvä vene kevyt
soutaa ja vaata se olla enempi intiaanikanootti mallinen kui savolainen
keskeltä melko tasapohjanen ja reuna nousta kuu kanootti suora ylös
lapsi ei saatua se kaatua ees kova yrittää neliheppanen kone perä ja
kaveri keula päristellä mennä liuku ihan puuosa lahota sitten huono pito
takia ja eve joutu eläke ranta#pusikko himputtaa tämä saattaapia herätä
kip vanha alumiini#soutu#vene kunnostus
```

In this stage, a list of words that are too common in the Finnish language so that they would distract the model was constructed. Such words include and, here, this etc. The full list of stopwords can be found from Github (www.github.com/tirri/thesis), where all the codes used in this thesis are also stored. The words in the stopwords' list were lemmatized with the same scripts as the messages.

From this point onwards, for the data handling and later for the modeling, a ready-made toolkit, a programming library called *Gensim* (Rehurek & Sojka 2010) was used. It is a well-documented, free to use code-library for topic modeling.

Most Social Science publications that were introduced earlier in this chapter use another code-library called *Mallet* (McCallum 2002) for topic modelling (an exception being Ahonen 2015, who used R-package called *topicmodels*). Mallet, as the name implies, is a general toolkit for textual machine learning. Gensim is specialised in topic modeling, and although it was introduced only 8 years ago, the introductory article (Rehurek & Soja 2010) has been cited by 1135 writers according to Google Scholar indexing. It greatly exceeds Mallet in the pace of development. In the version control platform called *Github* where the program developers share their improvements into the libraries, Gensim had 290 contributors, it was last updated 2 days ago and there were in total 3650 commits or changes to the code in the platform. The corresponding numbers for Mallet were 31 contributors, 631 commits, and last update was 12 days ago.[7]

With Gensim's *corpora*-module, the conversations were transformed into vectors. Each word in the corpus was given an id-number. Each document was described as list of id - number of occurrences -pairs. The result of this was a matrix of the corpus where each document in its vectorized form is its own row. At this stage, a dictionary binding the word-id's and the actual words, was also created.

In its matrix form, the most common and rare words were also removed from the corpus (see e.g. Ahonen 2015, 11). The result of this process was a corpus of 135 992 documents with 29 664 unique words. Finally, the data was ready to be fed into the model. (see eg. Wilson & Chew 2010).

Note on the bag-of-words -assumption

When individuals write on *Finland24*, one would think they are triggered to do so by something they have experienced or what someone else has written on the forum. Maybe they want to share information or a feeling, or get support from other writers. They open the computer or mobile device, choose a relevant conversation or start a new and type in in their personal style what they have in mind and press 'send'. The LDA-model, however, assumes the following process: Individuals have in mind a set of topics that they want to share with other users in Suomi24. For each of these topics, they have a bucket of words that represent the topic. For example, the bucket of the topic 'Illnesses and old age' might have words such as 'disabled', 'heart attack', 'doctor', 'medicine' and so on. Now, the writer puts her hand in the bucket and draws randomly

[7] The numbers were retrieved from Github on 29th August 2018.

a handful of words out of it. Writer then goes to their keyboard, types in the words without sorting them and publishes the result in the conversation tree.

The topic model assumes, the two methods', the 'human method' and the 'model method', outputs are equal. Needless to say that they are not, yet the topic models have proven to find relevant topics from different corpora (e.g. Levy & Franklin 2014). It seems to be, the topics we want to write about are revealed by the words we use, no matter how MANY *special* **linguistic means** we use to give the text rhythm and weight on the important parts.

The bag-of-words -assumption, or formally, the assumption that discrete data - be it textual or other form of categorical data - is exchangeable and can be represented as a mixture distribution, was a prerequisite for the development of the LDA-model (Blei, Ng & Jordan 2003, 994). For those for whom the order of the words matters, there are other computer assisted content analysis methods available (see e.g. Leydesdorff & Hellsten 2006).

4.2.2 Modeling the data: Latent dirichlet allocation

Machine learning algorithms, such as those of topic modeling, are typically designed in Computer Sciences, to answer one specific research question and to be used with one corpus. When applied in a different context, the model must be carefully chosen and adjusted for the purpose. (Yu, Kaufmann & Diermeier 2008.) For an unstructured corpus like the *Finland24*-posts, topic modeling has been shown to be well suitable (see e.g. Levy & Frankling 2014). Of the topic modeling methods there are, the original Latent dirichlet allocation (Blei, Ng & Jordan 2003) is the simplest and the most popular.

Latent dirichlet allocation (Blei, Ng, Jordan 2003) is a topic model that allocates probabilities for words and documents following Dirichlet-distributions, thus revealing the latent thematic structure of the corpus; hence the name. It is the simplest topic models (Blei 2012, 78) and as it has been proven to give valid results, it has become the most widely accepted topic model among the computational social scientists (see Purhonen & Toikka 2016, 15).

Being simple and popular were not the only reasons for selecting just this model among the many variations of topic modeling there are. It was also important that the model would suit the challenging data of the forum. The documents would be of very different lengths, and the number of issues a conversation concerns could be almost anything and vary from message to message. In LDA, it is allowed that documents contain varying amount of topics; as a matter of fact, all documents contain all topics but with varying probabilities (Blei 2012, 79).

Varying lengths of documents are taken care of by weighing them accordingly, so that a long conversation constituting of one clear topic would not dictate the whole contents of the topic. The liberty of topical distributions within documents is constructed using three levels of probability allocations. In many clustering models such as simple Dirichlet-multinomial clustering models, there are only two levels on which to operate: the corpus-level and the document-level. This restricts that all the words in a document are treated as a single cluster, i.e. belonging to one topic only, and any word can belong to only one topic. Instead, LDA is a three-level -model in which there are the corpus-level variables (defining the corpus-wide distributions of words and topics), the document-level variables (defining the distributions for each document) and the word-level variables (defining each word's probabilities for the corpus-wide topics and document-wide topics). Leaving the math aside, this allows for the conversations to consist of multiple topics. (Blei 2012.) Also, such a "mixed-membership –model" (term used e.g. in Lucas et al. 2015, 261) captures the context-dependent nature of the words. In different sentences, even within the same message, a word might be part of different topics (Blei, Ng, Jordan 2003).

The most obvious variables the model needs are corpus, dictionary and number of topics. Corpus defines, which document matrix is used for modeling. The corpus used in this thesis is the one described in the previous section. As the corpus in its matrix form does not hold any information about the words behind the numbers, a dictionary that maps the words' identity numbers to the actual words in natural language, is necessary. The dictionary was also constructed as described in the Data handling -section. The number of topics is the number of the hidden themes the researcher believes there are in the corpus.

Number of topics

The number of topics to find from the data with a topic model is a decision the researcher makes based on the study objectives, either with the help of another mathematical model, trials with different numbers of topics, theoretical knowledge or a combination of these (see e.g. Nelimarkka 2019, Nuortimo 2016, 30). The reasons for choosing the approach vary alike. Mohr and Bogdanov (2013) simply tell, "[a]fter exploring some alternatives, we settled on a 25-topic model" (ibid., 552). Ylä-Anttila (2018) explains that as his "study is a mixed-methods bent, I used both quantitative and qualitative methods in choosing the topic count as well" (ibid., 35). On the contrary, Nelimarkka (2019) recommend the use of a mathematical estimate because, the study shows, the subjective understanding of the best amount of topics varies and affects the results of the model.

There is no consensus as yet on how to find the right number. The best amount of topics is the amount that reveals all of the hidden thematic structure in the corpus yet compresses the information into easily interpretable word clusters. The problem is, how to find the equilibrium between these in a big data set. Representing each word in the corpus as its own topic won't do nor will only one topic over the whole, but there remains quite a scale between these two extremes. In the literature, the chosen amount of topics is argued to affected by the size of the corpus (e.g. Nuortimo 2016, Brauer & Dymitrow 2014), interpretability of the topics (Ylä-Anttila 2018) and/or topic coherence. Theoretical understanding of what the model is expected to find is, however, the key to the decision. For example, Purhonen and Toikka (2016) notice that different numbers of topics could answer very different research questions (ibid., 18). Within these frames, all studies tried different numbers and the exact amount was typically some "nice" number, divisible by 5 (see Table 1). The trial-and-error (see e.g. Ylä-Anttila, Eranti & Kukkonen 2018, 151 for a closer definition of this method) is advocated for a good reason, as the model is used in Social Sciences to find topics that humans understand as social phenomena. While there are computational methods for analysing the coherence of the topics, humans are still used as the gold standard of these evaluations. (See e.g. Levy & Franklin 2014 and Röder, Both & Hinneburg 2015, 2).

As there is not yet consensus what is the best way to choose the number of topics, the choice should be based on features of the method, study objective and data as is the case with any methodological question. In this thesis, the choice of using 200 topics is argued thus: The phenomenon of retirement contains possibly hundreds of sub-phenomena (see Chapter 2), and the study objective was to find fine nuances of them. A topic model of few topics would be too coarse. The large data (introduced in Chapter 3) could support possibly a very large number of topics (Nuortimo 2016, 30). Large amount of topics is also recommended by Ylä-Anttila, Eranto & Kukkonen (2018, 151), because if there appears to be minor, irrelevant topics, they can be abandoned later, but if the topics are too broad, they cannot be separated. Thus, the largest interpretable amount of topics was chosen to be used.

In the literature, the largest amount of topics found was 300 (Brauer & Dymitrow 2014), however their interpretation protocol is excessive and involves multiple researchers. The largest reasonably laborious amount of topics was 200 (Ylä-Anttila 2018), so this was the amount chosen to be used also in the context of this thesis. Why 200 topics was already at the limits of being too large, and why an even larger number of topics would have been too laborious will become evident in the following sections discussing the interpreting of the topics.

Running the model

As discussed in the section of data handling, the application used for modeling the data was built on a programming code library called Gensim (Rehurek & Sojka 2010). The LDA-module of Gensim is built on the original mathematical model of LDA (Blei, Ng & Jordan 2003), and it gives all the needed algorithms for accomplishing the modelling. The researcher only has to provide the algorithms with the aforementioned information of the corpus, dictionary and the number of topics (described earlier). Other parameters that the model uses guide the details of the formation of the topics. In Gensim, they have pre-set values that are strongly recommended, but they can be changed if the application of the model so requires. These 18 parameters inform the model about, for example, the (non-/)uniformity of the documents within the corpus and control the flow of allocating topics for documents. (See Appendix 1 for further information on the modeling process.)

4.3 Results

After the model had run for approximately half an hour, the iterations were ready and the program printed out the topics: 200 lists of words that frequently co-appear within the documents in the corpus.

The topics a topic model prints are not interpretable as such but they have to be labelled by human for further analyses (Grimmer & Stewart 2013). At this stage it was noticed, the model used in the thesis did not find the best fit for the data as nearly half of the topics were incomprehensible for the human eye. This is, however, sometimes the case with topic models and does not mean that the remaining topics were of poor quality (see e.g. Nuortimo 2016). The fit is usually adjusted by choosing different amounts of topics and analysing, which of the models forms the most coherent set. This was also done in the thesis, and fewer topics' models' were both mathematically and qualitatively analysed more coherent (see section 4.3.1). However, topic models are not necessarily used for finding the best fit but they are one more tool for understanding data and can only augment human reasoning (Griffiths & Steywers 2013). The task the model was set to do was not to find the optimal fit for the data, but to find as detailed summarisation of it as was possible. The model still provided 100 "good" topics based of which the qualitative analyses were, successfully, conducted (see Chapter 5).

4.3.1 Labelling and discarding the topics

The words the topic model had output (see Appendix 3) were first translated to English. The only reason for this was to make the process more transparent; the translated topics can be found in the Appendix 4 and the reader can validate the *labelling*, the names given to the topics. The labels were given based on the translations, yet the Finnish versions of the words were checked before the final decision on the label. The translations, as mentioned in the Introduction, are by the author.

The labelling of the topics is an extra step that there isn't in manual content analysis where the labels are given by the human-coders while they are coding (Grimmer & Stewart 2013.) Also, humans only find issues where there really is one, whereas topic models recognize co-occurrences of words, despite they might not have formed an issue, thus resulting in a topic like 140 which included the words *decision maker*, *generation, respect, Anne* and *modest* as the five most probable words. *Topic* model does not understand its reading; thus it cannot name nor recognize the non-coherent topics. There are other mathematical models that evaluate the coherence of the topics, however their usage is still disputed among the community using topic modeling. What they share is, they mimic the human understanding of the coherence within a set of words, which is why the best available method for interpreting the topics into labelled units is human evaluation. (Lau, Newman & Baldwin 2014, 531)

Grimmer & Stewart (2013) suggest that labelling should be done based on reading a sample of the documents with high probabilities of belonging to the topic in question. In this approach, the output of the topic model is relevant only when it provides direct access to original documents. The output is not viewed as such, instead, the focus is on qualitative thematic reading of the texts. (Ibid.) This method was used in e.g. Niemi (2016), where he would label most topics based on the most influential document that the topic was formed of. Some topics did not have one defining document but were more evenly formed by multiple key documents, in which case the most probable words (later also: keywords) in the topic were used to guide what the suitable label would be.

Opposite way of labelling the topics is to only use the keywords in the topics and decide the label based on the researcher's theoretical knowledge of the study objective. How many words are then used for the labelling varies: Purhonen and Toikka (2016) chose to label the topics by reading the 50 most probable words from each of their seven topics. Their data were presidential New Year's speeches, so they could categorize words like *battle* ('taistelu') and *common effort* ('yhteisvoim…') to be part of the coherent set of words under label *Ceremonial Speech.* Similar approach was used in

Ylä-Anttila (2018), where 20 most probable words from each of the 200 topics were scrutinized and given label or discarded. A notably different application of this strategy was used in Ahonen (2015), where the possible categories - legislative regimes - were known beforehand.

The third way to approach labelling is to combine the two abovementioned: first to look at the most common words in the topic's output and give an initial label, then to find the documents that were most influential to the formation of that topic and validate by qualitative analysis if the given label is suitable. This combination was used excessively in Brauer and Dymitrow (2014) where they even created coding matrices in order to find the relevant topics out of the so-called 'noise' and to avoid subjective interpretations of the labels (ibid. 36). In e.g. Ylisiurua (2017) some key topics would have been abandoned from the analysis were the close reading of the documents been skipped. Also Ylä-Anttila (2018) validated the labels given by keywords by reading the ten most influential documents for each of the non-abandoned 17 topics, and abandoned further three topics.

Labelling based only on topics' keywords has shown to provide good results (e.g. Purhonen & Toikka 2016, Ylä-Anttila 2018) and it is less laborious than close reading the messages. Thus, in this thesis, a special effort was put onto developing the labelling system. For none of the studies using topic modeling described how exactly did they decide over the right label nor if the topic should be discarded; Nuortimo (2016) describes he gave a label if the keywords formed a set that made *"sense in the light of the author's knowledge on the subject"* (ibid., 38). Looking at some of these topics, it is likely that another researcher would have judged their coherence differently (see e.g. Purhonen & Toikka 2016, 19). Also, there is great variation on how many of the keywords are used in labelling, the scale found being at least from 10 (Ylä-Anttila, Eranti & Kukkonen 2018) to 100 (DiMaggio, Nag & Blei 2013). In this thesis, 15 most probable words were used for this purpose.

The intuitive labelling was first tried in this thesis' context. This turned out to be problematic: many topics seemed to consist of multiple topics, or the first most common words in them formed a sentence-like combination that was easy to misinterpret. Interpreting the topics under labels systematically is very rare, or at least it has not been reported (see, however, Mimno et al. 2011). This was found problematic for the replicability of the analysis, which is why a methodological approach for this task is suggested and used in the current thesis. The methodology is presented below and is based on the meta-analysis of the labelling systems of other writers and the authors' intuitive understanding on how the process could be made replicable.

Rules for labelling

1. Recognising a coherent set of words should be intuitive for a human (Mimno et al. 2011). Thus, it was reasoned, the connection could be found by looking at only the five most probable words. If it was possible to name a topic based on just the 5 most probable words, this was done. Such topic was, for example, the topic number 33, where the five most probable words were *well-being, economy, workforce, welfare state* and *employment.* This topic was obviously about welfare state economics: all the five most common words were directly related to it. So the topic was labelled as *Welfare state economics* without scrutinizing the rest of the words (that did not oppose the label, being *"challenge", "to turn", "-building", "entity", "homelandic", "simultaneously", "model", "return", "this day"* and *"one in working age*").

2. Otherwise, if the five first words in a topic were "promising", as if some of them might be about the same issue, I would look at all the 15 most probable words, and categorize them into three subsets: *directly related, common in the context* and *not related. Directly related* were the words that were recognized as belonging to an issue x. *Common in the context* -subset included words that were commonly used in the context of the issue x, even if they were not directly about that issue. Words in the *Not related* -subset included, as the name hints, the words that were not related to the issue x. This categorisation was inspired by the other topic modelling studies' topics: It is common that the words in the topics can be thus divided, even though this is not stated in the article (see e.g. Purhonen & Toikka 2016). Mimno et al. (2011) consider this type of a topic to be *unbalanced*, including it into a group where there is a fair chance another scholar would not have categorised it as a topic, thus these topics were given special attention. If more than half of the words were either categorised as "directly related" or "common in the context", I would decide the topic was worth the label. If not, I discarded the group of words as a non-coherent topic.

As an example of a topic where the rule number one did not suffice for giving a label to the topic but that was "promising" as described in rule number two was the topic number 134. The topic's five most common words were congregation, church, spiritual, lawspeaker and pocket. The four first were promising, but pocket is not related to them in any way. The rest of the fifteen most probable words in the topic – general, first congregation, natural, bishop, member, to be built of, Kuopio (i.e. a city in Finland), proverb, excursion and diligent – were needed before I could be certain there really was a topic that made sense as a coherent set of words. So I categorised the words: Six (namely: congregation, church, spiritual, lawspeaker, first congregation and bishop)

were directly related to church's institutions, four (member, to be built of, excursion and diligent) were "common in the context", that is, are general words that are very common within church context. The rest are not related to the church, but as 10 of 15 makes more than half, I decided the topic number 134 was about an issue and deserved a label. Thus, I named the topic "Religious institutions".

In the case of topic number 134 or Religious institutions, it was quite clear there was a topic after all. Some topics were more ambivalent, like topic number 7. I classified its words as follows: directly related was medicine; common in the context were to eat, to order, to use, to quit, to feed, background, according to, to drift and suicide; and not related were perfection, to hit, light, the size of and pass. As can be seen, only one of the words I labelled as directly related, forming its own label. Yet eight words in the topic were very common in medicine's context, so it looked like a maybe-topic. Only after using my rule of classifying each word, I could be certain it was a medicine-related topic: 9 out of 15 words were either directly related or common in the context, thus more than half.

Another way the topic confused me was when there seemed to be two or more coherent sets of words within one topic, similarly to what Mohr & Bogdanov found (2013, 552-554). One such was topic number 55, whose words I classified in four groups. In the first group were words relating to family members: *mother, father* and *brother*. The second group consisted of words of ageing process: *to die*, *birth* and *old*. The third group was about negative feelings, including words *sad* and *to be disappointed*, and to the fourth group I collected the rest of the words that seemed not to be related to any of the above nor formed their own group: *he who brings, truly, along, building, music, effort* and *to warn.* I was unsure if this really was a topic or not, but I reasoned it was, as both negative feelings and dying, birth and oldness are occurring in family setting. Moreover, negative feelings emerge when someone in the family dies. So my human intuition told me, there is a topic, even if it did not fit to the rules I had created. Thus, I added a third rule to my decision making system:

3. If, within the fifteen most probable words in a topic, there were multiple coherent subsets of at least two words (like was the case with the topic number 55: family members, ageing process and negative feelings each had at least two words that held a distinct meaning within the subset), such that each subset is "common in context" to the other subsets within the topic (again, as I reasoned in the previous section, the three subsets in topic number 55 commonly co-occur), and the subsets together form at least half of the words within the topic, then the topic will be what I call a "combination topic": a topic combined of multiple themes.

With the rule number three, I could now label the topic number 55 *as Family, birth and death and negative emotions*.

Of the 200 topics the model printed, 103 were given labels according to these rules. For the sake of conciseness, the complete list of the labelled topics is not presented here but can be found in the Appendix 4.

Abandoning topics and computationally measured topic coherence

Approximately half of the topics were, thus, such that they did not form a coherent set to the eyes of a human reader (classified as *random* topics in Mimno et al. 2011). An example of a non-topic was the topic number 64 which had the words *border, countryside, to miss, salt, necessary, hall, cloud, quite a [something], well, gay, nest, respect, ball, to protect* and *to drown* as the most common 15 words. With good imagination, some of the words can be seen as belonging to a common topic: in cooking, salt is necessary and might be missed; or a nest should be protected, but there really is no clear "common factor" for the words. These kind of *intruded* (Mimno et al. 2011, 264) "topics" were disregarded of the analysis.

Also the computational analysis of UMass topic coherence (see e.g. Stevens et al. 2012) argued the 200 topics' model did not find an optimal fit for the data. In fact, when compared with the models of other numbers of topics, the 200 topics' model fared the worst: The 200 topics' model had the lowest topic coherence score of -3.69; 100 topics' model was already better with the result of -3.40; 50 and 25 topics' models were already very good with the scores of -2.52 and -2.64 (respectively); and the 5 topics' model had the best score of -1.76.

In this study, the modeling was done using 200 topics. With this large a number, it is not atypical that many, if not most of the topics are irrelevant for the study (see e.g. Ylä-Anttila 2018, 10). This is because the model has not the understanding of its reading as a human would have. It only recognizes that words do co-appear frequently, which may happen out of coincidence, or because of the specific language or a single writer, or because some or multiple documents in the corpus somewhat biased the results. But, if the model has a good fit in the data, the topics should be *coherent* and only abandoned because they are *not relevant* for the research questions of the study (ibid.), however in this study half of the topics were classified non-coherent by the human-analyser and the computational topic coherence score remained low. It is argued, though, that topic coherence for all the topics was not the goal of this study, instead, it was to find a detailed enough summary of the themes in the corpus for spotting the conversations where retirement narratives would show up. Also Nuortimo (2016) did

not find it problematic that only 29 of his 100 topics were considered coherent and taken into further analyses.

At this point in the current thesis, every topic that formed a coherent set as described above was kept in the analysis, regardless of if I considered it to be related to retirement or not. Later, most of the topics were considered to be general topics, meaning they describe issues that could appear in any kind of conversation. This process of abandoning the irrelevant yet labelled topics was also designed by the author as there were no examples of how this could have been conducted in previous literature.

4.3.2 Interpreting and validating the results: Which topics are about retirement discussion?

Ylä-Anttila, Eranto & Kukkonen (2018, 155) suggest the data may be limited to be about one theme by using keywords. In this study, it was hypothesised, the keyword 'eläke' (pension or retirement) would limit the data into the discussions that concerned the issue. A major concern when choosing the data was that retirement is such a general issue that mentioning it might not mean the conversation really was about retirement. However, it was logically concluded that it is hard to discuss an issue without mentioning the name of the issue, so when discussions about retirement are taking place, the word will be mentioned. Yet, retirement is such a general word it may be thrown into conversation that was about something completely different issue, so the logic does not work exclusively. Looking at the Appendix 4, it is clear that many, if not most, topics in this study, were such that they could have come up from the forum without the restriction to retirement related messages. It had to be validated that mentioning retirement would yield also specifically retirement related themes.

To find these, specifically retirement-related themes, a control group was used. The control group was a random sample the size of 60 000 messages of all the messages in the Finland24-forum without the restriction to retirement related words. This sample is called the *general discussion*, as distinguished from the *retirement discussion*, referring to the messages with a retirement related word. If the control group discussions would output different topics than the retirement related intersection of it, and if the different topics were theoretically relatable to retirement, it would be validated that the latter conversations truly were those of retirement. The process would also give response to the research question, what were the topics of retirement people discuss on the forum.

The sample of general discussion was modelled with the same 200-topics model that I had used for retrieving the retirement-related topics. Thus, I had another list of 200

topics, but this time, the list items should be topics of general, not retirement specific, discussion. After labelling also the general discussion's topics, the two lists were compared in order to find overlapping and different topics.

Each topic in each of the lists were compared with each other. If the topic was the same, it was considered as 'a general topic', meaning it discussed an issue that could have come up in any kind of a conversation. If there was a topic that corresponded to none of the other list's topics, it was considered unique to retirement discussion. The labels of the two groups did not have to be equal to be considered the same, but their issue should be very close. Thus, e.g. the topics *Meals* (in general) and *Eating* (in retirement specific topics) were marked as corresponding. As a result, there were 78 topics that were prevalent only in retirement specific topics and, as a curiosity, 71 topics that were present only in the general topics list.

Thematic grouping of the topics

The 78 distinct topics among (hypothesized) retirement specific discussion showed, the two corpora yielded different topics. But these topics were still very specific and were thought not revealing if the differences were because of the other set was about retirement related topics or simply because it is likely that a topic model always gives slightly different topics. To compare if the differences were due to retirement related conversation taking place, a qualitative comparison was conducted.

The topics of both (hypothesized) retirement specific and general discussion's topics were summarised under wider themes that emerged from reading through the topic lists. Theoretical understanding of retirement was used for finding the suitable themes from the (hypothesized) retirement specific topics, as this was recommended by Brauer and Dymitrow (2014): the topics should be categorized so that the results are "most relevant to the aim of the research, most notably by anchoring it in specialist literature." This way, they state, there is a minimum risk of the subjectivity bias. (Ibid., 35.)

An example of such a theme was *Regulation* that summarized the general discussions' topics *Bureaucracy, Judgement* and *Law and regulations.* The complete documentation of these comparisons is presented in Appendix 5. The following section discusses the themes and how they indicate that however general theme retirement is, discussions of it still held distinct thematic structure which the topic model had identified.

Comparing the themes and topics between general and retirement related discussion
Both the retirement specific topics and the general topics formed 15 themes of which 5 were unique to it. Ten of the themes were shared between the two groups:

Table 2: Similar and different themes in retirement related and general discussion.

Retirement discussion's themes	**General discussion's themes**	**Common themes**
Social Issues	Crimes	Media
Social Security System	Weather	Money
Social Development	House and living in a place	Politics
Retirement transition	Technology	Meta
Life and Feelings	Transportation and Vehicles	Regulation
		Religion
		Relationships and Family
		Spare Time
		Work and Livelihood
		Illness/Body and Illness

As was expected, most of the themes found were the same between (hypothesized) retirement related and general conversations. The common themes are of such issues that are, by a quick glance, very typical to arise in any conversation. To analyse the common themes in more detail is a subject of another study. Here, the focus is on the five themes that were specific to (hypothesized) retirement discussion.

As was discussed in the Chapter Two, the four most occurring issues in retirement research of today are retirement exits, older workers, pensions, and the retired life (Ekerdt 2010). If the themes that were classified as retirement specific were about retirement, it may well be said the theme *Life and Feelings* corresponds with the retired life, that *Retirement Transition* is a synonym to retirement exits, and that pensions are part of the *Social Security System*. That the themes *Social issues* and *Social development* also emerged as specific to discussion on retirement is less evident but can be reasoned to be related to the nature of retirement. Retirement system is a major theme when the future of society is being discussed, which would explain the theme *Social development.* Likewise, retirement is a specific phase in life, typically associated with decreasing resources and increasing vulnerabilities, thus relating retirement with other social issues. But the reasons why the themes showed could as well be due to some

particular feature of the forum, undermining the results. To analyse if these promising results were 'true', randomly chosen discussions where these themes' topics were prevalent, were read. The discussions under the themes relate to the themes with a good level of accuracy.

After all these analyses the result of the topic modeling is, the themes that people discuss on the forum around retirement are *Social Issues, Social Security System, Social Development, Retirement transition* and *Life and Feelings.*

Table 3: Retirement related themes in Finland-24

Retirement transition	Social development	Social issues	Social security system	Life and Feelings
122: Getting older and out of working life (of age, retirement age, to turn, age, to get older, to be born; to rise, work, working life; earnings-dependant, daily benefit, retirement pipe; to hold, to get, to stay)	165: Social development (union, to develop, citizen, development, societal, international, culture, social; research, to present, to show, source, level, to give birth, to aim)	20: Socio-economic groups (poor, rich, money, retiree, unemployed, poverty, student, family with children; National Coalition party, Finland, people, to live, human, common; to hold)	136: Schools (school, student, teacher, to shout, to teach)	154: Interpretations of experiences (to experience, attitude, to face, interpretation, atmosphere, experience, mentally, mental, best, picture, violence, valuable, exam, sensor, consequence)
74: Age, retirement and activities (young, old, retiree/pensioner, wrinkle; to get into, to do, to stay, work, to begin, gang, to know how, to have to, to run to be saved, best)	171: Development in the Finnish society (unemployment, labour force; increase, to increase, to grow, amount, number; problem, future, country, Finland, current; immigrant, Finnish; to need)	10: Badness and prison (black, death, prison, to lie, to dare, bottom, castle (in Finnish slang, used of prison), prisoner; little, to fly, meter, safe, old age, surface, spiritualize)	141: Pension system (retirement age, pension fund, pension system; increase, increasing, current, to rise, ageing, population; bed, long, to hold, spare, channel, to do)	183: Human life (life, human, to live, to do, to think)
	180: Social classes and ideals (merchant, labourer, communist, socialism, capitalist)	167: Drugs and society (alcohol, drug, illness, health, spirit, into drugs to cause, usage, cost, surveillance; energy resources, social, Ministry of health, Ministry of Justice, to interpret)	142: Elderly care (old person, to nurse, treatment, nurse, to keep)	45: Happiness (to be satisfied, joy, happy; home, sun, life, scene, eternal, freedom; step, around, piece, me, to raise, to spend)
	35: Frustration over system called society (system, human, indifference, to be enough, arse; workplace, work, to employ, social security office, money, to have to, to get to, away, to take; to do)	192: Stigmatisation (citizen, name, feeling, victim, gift, human, smart, to feel, to be ashamed, to stigmatize, feature, gossip, thinking, cow, shame)	199: Taking care of children and the elderly (child, family, old, home, care taking, of age, others, to take care of, elderly, age, fatherless, years old, upbringing, to grow, alimony)	9: Goodness (love, peace, perfect, fellow human being, security, soul, power of change; individual, to renew, home; warning, free/unfastened, betrayal, land, fatherland)

	39: Society and development planning (human, society, world; consequence, reality, problem, future, to solve, to lead; to hold/like, on offer, to do , to live, to begin, mercy)	184: Population statistics (minor, population, statistics, comparison, Ministry of Employment)	143: Social benefits (month, income, housing allowance, income support, benefit)	
		159: Crimes (police, crime, to investigate, illegality, to accuse)	152: Basic social income support (to raise, purchasing power, person with low income, increase, social security, livelihood, standard of living, labour market support, poverty, tax-free, basic security, to raise, basic part, income, basic daily support)	
			176: Benefits (benefit, pension income, statutory, automatically, individual, to suspend, to avoid, to perform, possessing, task, to define, supply, watch, clearly, finally)	
			99: Applying and the benefits for the disabled (to apply for, to allow, decision, applicant, application, disability retirement, Kela*, disability unemployment, rehabilitation, rehabilitee allowance, working ability, temporary, scientific, basing, sii**)	
			57: Gates in handling an application from a person's perspective (to wait, to let, hurry, entrance, to get in, extension positive, retirement decision, actively, vihta***, voice, to choose, fear, poor, fresh)	

*(The Social Insurance Institution of Finland)
**not a word
*** Traditional Finnish massage equipment

4.4. Reflections on the method

As discussed in the beginning of this chapter, topic modeling as a rather new method has not yet established best research practices within Social Sciences. It is a sensitive method where subjective decisions, like the number of topics and topic labels, still have a great impact on the study outcomes. Most of the scholars using it have, thus, given special attention for unifying its applications by developing some aspect of the model, or at least explaining their reasoning behind the choices. Following this tradition, the following sections will discuss if the 200 topics' model was the optimal choice, as well as analyse if the systematic labelling yielded correct labels.

4.4.1 Too many topics?

The large amount of topics that was used was argued to provide the thesis with a richer, more holistic view over the data than what a smaller amount of topics would have. The decision resulted in making the labelling of the topics highly laborious, due to which only 15 most probable words from each topic were used in the process. Other topic modeling papers have typically used less topics and more words and even documents in interpreting them (see e.g. Ylä-Anttila, Eranto & Kukkonen 2018), which allows the researcher to skip more words from the analysis of their coherence and might thus help the labelling.

During testing the validity of the model, the data was tested with 5, 25, 50 and 100 topic's models. Mathematical topic coherence analysis observed the coherence score worsened by increasing the amount of topics. Thus, the only base for choosing the 200 topics' model lied in the theoretical understanding of the research objective: only fine granulation of the topics would find the detailed topics describing retirement transition (Nuortimo 2016). This was tested by comparing the labelled topics of the other models with the retirement specific themes of the 200 topics' model. This way it was possible to see if the themes found were consistent, regardless of the amount of topics used. The table below summarises the results:

Table 4: Retirement specific themes' prevalence in different models

	5 topics model	25 topics model	50 topics model	100 topics model
Social Issues	No	Yes	Yes	Yes
Retirement transition	No	No	No	No

Social development	Yes	Yes	Yes	Yes
Social security system	No	Yes	Yes	Yes
Life and Feelings	Yes	Yes	Yes	Yes

The results show that all themes, except for *Retirement transition*, were present in nearly every other model. It was to be expected that the 5-topics model would not find every theme; it is unlikely that only the retirement specific themes would have ended up into its results as there were so many themes that were of general discussion also in the 200-topics model. What was interesting was, not 5, 25 nor 100 topics' models could find the theme *Retirement transition*, which was later selected as the most intriguing theme and analysed in more detail. It was interpreted, retirement transition was a very subtle topic within the corpus and needed very fine granulation of the corpus' thematic structure in order to be found (see e.g. Nuortimo 2016, 30). This finding suggests, that even more that 200 topics could have been optimal for finding topics related to *Retirement transition.* But, as already discussed, a greater number of topics would have been too laborious to analyse.

4.4.2 Is the labelling system providing the right labels?

As is the case with machine learning models, the workings of the topic model are a so called black box. The researcher puts their data in, and out come the results. What happens in between is a complex process of calculations based on probability distributions that are known in theory but it would be extremely laborious to go each step of the calculations through. (see e.g. Why are… 2017.) The mysteries this leaves the researchers with when they are investigating the results clarified in the context of this study when the conversations were chosen for the analysis.

The reading of the messages was necessary also in order to verify that the labels for the topics and themes were correct. Three conversations that held the highest probabilities of belonging to the two topics, *Getting older and out of working life* and *Age, retirement and activities*, in the theme *Retirement Transition* were read. Respectively, these conversations had the probabilities of 0.70, 0.68 and 0.62, and 0.63, 0.58 and 0.55 to belong to the topic. I compared the conversations' lemmatised version to the fifteen most probable words in each topic and counted their frequencies among all the words. The results were curious. The topic number 122 pulled the probabilities

of 0.15, 0.26 and 0.12, but the topic number 74 only held the frequencies of 0.04, 0.06 and 0.09. Qualitative reading of the messages revealed, the posts in the topic number 122 (Getting older and out of working life) were all such that wondered how the pensions accumulate during unemployment and what benefits do they get if they become unemployed at some given age. So the high probabilities of the fifteen most common words were rightful, both confirming that the label given to the topic was appropriate and that the model had been able to recognise a clearly distinct topic.

The result that the conversations that were supposed to be highly relevant for the topic number 74 (Age, retirement and activities) did not seem to be so for the human reader was more worrying. Reading the messages was even more puzzling: they were about age, retirement and activities after all! The first of these conversations was planning to continue something starting on retirement, the second wanted to gather people for pushing others to retirement, and the last asked where they could find an association for pensioners and gave details on what the association should be doing. How was it possible that the model recognised these conversations belonging to a topic, when they did not even contain the words that the topic consisted of? Even going through 100 most probable words in the topic did not solve the mystery. Somehow, the model was making a good job in its black box, although a high probability of the conversation to belong to the topic did not directly correlate with the relevance of the conversation to the human interpretation of the topic.

Another confirmation that labelling and thematic grouping of the topics was working as it should came from comparing the interpreted themes of the general discussion's model with the sub forum themes as listed in Finland24 (see Table 5 below).

Table 5: Comparison of the general discussion themes to the main topics in the forum.

Themes found by grouping the topic model's output	**Main topics as listed in the forum**
Money	Economics
Politics; Crimes; Religion; Media	Society
Relationships and Family	Family; Relationships; Sex; The young
Spare Time	Spare Time; Hobbies; Pets; Travelling; Food and Drink; Entertainment and Culture
Work and Livelihood	Work and Studying
Illness/Body and Illness	Health; Fashion and beauty; Sports and Exercising
House and living in a place	Home and renovation; Municipalities

Technology; Weather	Science and Technology
Transportation and Vehicles	Transportation and Vehicles
Meta	
Regulation	
	Groups

If the topic the model found was not among the main topics in the forum, it was compared to the subtopics of the forum. By using this method, nearly all the themes that were found by thematically grouping the topic model's output, found their counterpart from the forum. This is the best available validation the model functions as it should.

5 Qualitative analyses

5.1 Selecting the relevant conversations

In the previous chapter, the themes *Social Issues, Social Security System, Social Development, Retirement transition* and *Life and Feelings* were found to be endogenous to the retirement discussion in the Finland24-forum. If retirement narratives existed within the conversations, they should be found within the conversations in these themes. Looking at the themes it was obvious that *Retirement transition* would be the optimal theme for the narratives: as discussed in Chapter 2, retirement narratives are stories about the norms that guide late career decision-making (e.g. Kujala 2006) and post-career life expectations (e.g. Smith & Dougherty 2012, Frantsi 2012).

There were only two topics in the theme, *Getting older and out of working life* and *Age, retirement and activities*. From both of the topics, 14 conversations, or 28 in total, were selected for analysing them qualitatively.

The amount of conversations analysed qualitatively is much smaller than e.g. Ylä-Anttila (2018) who used all the posts in the chosen themes (or *frames*), nearly 1,500 in total, for his qualitative analysis. This large an amount was impossible to analyse within the time limits of this thesis. Instead, special effort was put forth to select only the most relevant conversations for the analysis:

In order to be chosen for the analysis, the conversation had to be considered by the model to belong into the topic with at least 5 percentages probability. Besides that, it had to contain at least 5 of the 15 most common words in the topic, which was considered a restrictive enough a rule for humans to agree with the model that the conversation really is about the given topic (see e.g. Ylisiurua 2017). Also, the number of words in a conversation was limited to no more than 300 because it is more probable for longer conversations to satisfy the criteria of five words match. Of such conversations, a random sample of 28 messages or message chains was taken. The descriptions of the conversations can be found from the Appendix 6.

5.2 Conduct of the analysis

How does one recognise a norm or a narrative in a discussion forum? Referring to the second chapter, a norm is an unquestioned ’truth’ that guides individuals’ behaviour in society. If it is questioned, the claim needs solid argumentation. Narration means spreading knowledge other than scientific, and cultural narratives are ’true stories’.

It appeared to be extremely difficult to spot norms and narratives straight from the conversations even with these definitions. Narratives that Kujala (2006), Frantsi (2012) and others had found could be used as guiding lines, but it was expected the narratives on the forum would somewhat differ from them. So, instead of directly coding the retirement narratives and norms, the discussions were first structured with descriptive questions. Norms and narratives were later absorbed from these descriptions. (Eskola & Suoranta 1998.)

However, it was not even known beforehand if there were norms and narratives to be found in the forum nor what they would be like. What was known was, if there were narration of retirement or a norm related to it, that part of the conversation should mention retirement or refer to it otherwise. Based on the third chapter where the conversations that mentioned retirement were acquainted with, it was evident there would be large sections within the discussions that had nothing to do with retirement or pension. Thus, the primary stage of coding the messages was to separate the parts that did not address retirement from those that did. The non-analysable parts were given the code *Not related to retirement*. In some cases, those parts covered large parts of the conversations. Most of the data was, however, related to retirement.

The messages were coded with the same Atlas.ti program that was used when getting to know to the messages. The relevant phrases that surrounded an occurrence of the word 'eläke' (retirement or pension in Finnish) were marked as a quotation and given at least two codes on the first round of reading the conversations. The first code answered the question, **in which context was retirement mentioned in the writing**. This question was posed in order to find the places in social discourse where retirement would be mentioned. The second code was given to describe, **how retirement was mentioned** and thus indicate if it spread a narrative or scientific knowledge. Some quotations received multiple instances of these code-types if the quotation represented multiple contexts for mentioning retirement or multiple ways for doing this. Later rounds of reading and coding answered more specifically to questions that arose from the topics the discussions were supposed to represent: **how does the conversation discuss *Age, retirement and activities* or *Getting older and out of working life*?** It was also specifically asked, **is there a retirement narrative taking place in here** (which formed the main theme *Narrative*) and **what is the argumentation like** (resulting to the separate theme *Argumentation* that will be analysed in section 5.3.1).

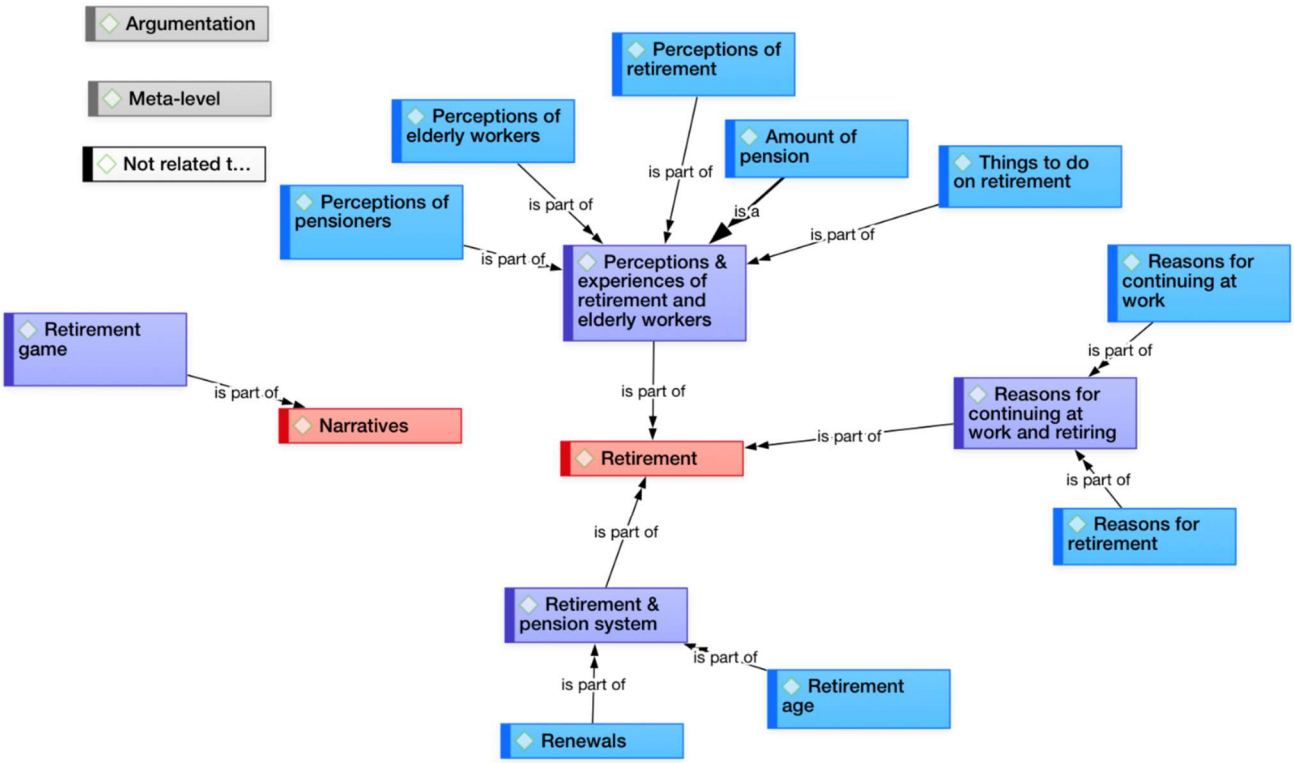
The *Code forest* and *Network* –tools embedded in the Atlas.ti-program were next used to group the codes hierarchically. Each code that was linked to a quote was linked under a superior, more abstract code. The superior codes were linked to even more general abstractions of the phenomena the codes represented. The groups were not set

beforehand but emerged from the codes. For example, a theme of *Things to do on retirement* was formed from all the codes that discussed the expectations and stories of what people on retirement were doing. Together with four other such themes, the theme was further summarised under the more general theme of *Perceptions and experiences of retirement and older workers*.

Once the hierarchy seemed final, analysing the conversations was started. Initially, each code-group was separately analysed, asking, what does this group tell about the writers' perceptions of the transition to retirement? These perceptions are collected and compared with the narratives found in previous studies in the section 5.3.2.

The coding and their grouping were iterative with writing the analysis: when an interesting phenomenon emerged, the conversations were read again in search for other instances of it. The most remarkable of the results of this process is the found narrative of retirement as the final line it the career race, described in section 5.3.3. After the analysis was ready, there were 147 different codes related to 156 quotations, categorised under the abstract codes represented in Figure 4:

Figure 4: Higher level codes in Atlas.ti's Network-view.



5.3 Findings and their interpretations

5.3.1 Argumentation in the forum: science or narratives?

In the second chapter, metanarratives were mentioned to be opposite to the scientific form of knowledge, so the first clue if there were narratives on the forum was to be found from analysing how the writers base their arguments. If the argumentation is based on peer-reviewed facts, the knowledge that is spread on the forum is not narratives but scientific (yet the scientific form of knowledge may lead to the norms of retirement). If, instead, it is based on vaguer sources or on no sources at all, narration can be said taking place (Lyotard 1984, Berger & Luckmann 1991).

It was to be expected that the argumentation in the forum is often based on the writers' opinions because it is a discussion forum where people are free to speak as they wish, not, for example a scientific conference where all argumentation should be based on previous research. There were, however, some writers who did not share a narrative but science based knowledge in the forum. The most remarkable of these were the "Old age retirement wizard" who, as can be read from their profile, are the pension experts from ETK. They answered to the questions people had about their pension with long expert explanations.

Besides the "retirement wizard", there were other writers too who used references to support their argumentation. They referred to different media sources, from videos and blogs to news articles. The quality of the sources varied, *Seura*, a yellow press magazine with the main goal to amuse people, was taken as seriously as statistics in *Taloussanomat*, the largest business online daily newspaper in Helsinki. The former I view as spreading some narrative, the latter as scientific form of knowledge.

Much more common were the arguments that were based on sources emerging from one's own life. Such included common knowledge, own experience and intuition. Then there were the arguments that did not seem to need no base at all, such as clever plays with words, passionate and emotional writings and pure claims indicating this is how things are, end of story. Thus, the type of argumentation indicates that narratives are taking place on the forum as was expected: Narratives and traces of them occur in every-day speeches and actions, and they are bound into the structures of the society where they are being told. Master narratives' echoes can be heard anywhere in the social discourse, also in the discussion forum.

5.3.2 Transition to retirement

Synonyms for the act of retirement

When finding out what were the perceptions towards retirement on the forum, it was first seen what kinds of terms people use of retirement in the conversations selected for the analysis. In the literature, retirement was described as a positive transition, an expected and accepted leisure-phase in life (e.g. Adams & Beehr 2003), but it was also noticed that for some, retirement was forced by a norm (Kujala 2006).

All the different ways to speak about retirement were there in the forum: To retire, (*eläköityä* in Finnish) was the neutral term to describe the end of career and beginning of a new life phase with pension replacing salary as the main source of income. To withdraw (*jäädä*) and to move on (*siirtyä*) were also classified as neutrally describing retirement. For some, retirement was something to get into (*päästä*) and pension was received (*saada*) and earned (*ansaita*), but there were also those who had to (*joutua*), were moved to (*siirtää*) or became kicked off (*tulla potkituksi*) on to retirement.

Retirement choices and the norm of retirement age

By far, the most common reason, or part of the reasons, given for plans for or actualised retirement on the forum was pension eligibility age (in 14 conversations). Similar to what previous research has found, most messages intermediated acceptance if not expectation of age as The Reason for retirement. This obviousness of retirement is called here *the norm of retirement,* and the term *norm of retirement at pension eligibility age* refers to the normative expectation of scheduling retirement at pension eligibility age. (Solem, Syse, Furunes et al. 2016.)

Planning for retirement took place in seven conversations where the norm of retirement at pension eligibility age showed clearly: No commentator questioned the retirement plans of writers closing to their 65th birthday, and indeed, when planning for retirement, there was only one message where pension eligibility age – take or give a couple of years – was not the scheduled timing of retirement. A taxi driver whose industry was no longer profitable did not mention their age; their reason for retirement was problems in business. Anywhere else in the forum, concerns of reaching pension eligibility age at work guided the plans to retire. The writers were asking for advice about the best exact timing of retirement in order to maximise pension: “If I get to income-dependent unemployment benefit ... until I am 65 years old, will I lose anything [of my pension] ...?”. Two different commentators who were able to achieve their pension eligibility age still wondered if they could add up to their pension by working a little longer:“... if I work let’s say until I’m 65? Will the pension accrual rate fall back to 1.5 %?” and “can I in my situation have the increase for late retirement if I’ll stay on the

extra days until 65 [years of age]?". 65 years, the pension eligibility age, is repeated (see also Takala et al. 2015). Another writer expected to "have a pleasing job hopefully until retirement age" – but no longer than that. Yet another who was about to change profession argued the choice through retirement age: "I cannot continue in my current job until retirement, so I have to prepare for changing profession". What they meant was that they were going to have to retire from their current job quite early, but they would not be allowed to have pension then. The re-education they planned was instead supposed to provide them with a job until 65 years of age, or whatever the pension eligibility age would be when they reached it.

Because old age pension and disability pension were not separated in the analysis, illness was not surprisingly the second most common reason (in 4 conversations) for receiving pension. Also in the literature, most significant predictor of early retirement has been found to be (perceived and actual) decreased ability to work, which is mainly determined by health and working conditions (e.g. Kyyrä & Paukkeri 2015, Elovainio et al.2003). A sad example of this will be described in the following sections in the form of a 28-years-old writer, already unable to work.

Layoffs and non-profitable business were also given as reasons for retirement plans in single messages (see also Tuominen & Gould 2010, von Bonsdorff et al. 2010), but also a successful business transaction had allowed one entrepreneur to 'retire', or rather stop working, early ahead of their pension eligibility age. This reflects the possibility of the well-off to choose from the alternative narratives of retirement (Van Sluytman & Torres 2014, Frantsi 2012).

Moralising the older workers

In previous studies, older workers have been found to be competing against contradictory expectations that originate from their colleagues (see e.g. Kujala 2006) and from the political narratives (see e.g. Taylor & Earl 2016). Also in the forum the writers commented and eagerly moralised others' retirement, part-time retirement or non-retirement choices (Ekerdt 2010, Adams & Beehr 2003).

There were two types of older workers mentioned in the forum: those approaching their retirement age and those already retired who were receiving pension while still in paid work. The moral norms concerning the latter group were completely missing from the literature of retirement narratives, which made these perceptions especially interesting.

Keeping a full-time job and receiving full pension benefit at the same time was mainly seen negatively in the forum. Pensioners' reasons for continuing at work were questioned: "when you are used to receive your pay check for barely chilling out in your

work place, you don't want to stay away even if you retired". The norm of retirement at pension eligibility age manifested in many comments. It was criticised that people in their 60's should continue working when the jobs could be given to the young who have it hard to find any job.

As one writer put it, "… these guys get pension every month and salary on top of it ... if you collect pension, you should not keep your work". The most extreme opinion was,

> *"all who have reached retirement age, they should be denied any kind of work in any firm. If nothing else will do, they must be carried out of the work place. And they should be forbidden working when they received pension and given fine if they did".*

Despite the previous writer is against all kind of work for pensioners, part-time work during retirement was not judged elsewhere in the data. A bricklayer, "a not so young pensioner", had done the masonry for a writer with a very low cost, which was appreciated (although, as an online forum's dismissing culture demands, the bricklayer was mocked for being a farmer too, still receiving agricultural subsidies, which was considered as the reason for being able to charge the lower price). There was even one pensioner who themselves commented they were having a part-time job to get some extra money, which was not commented in any way.

The older workers who were yet but approaching retirement age were faced with equally contradictory attitudes than those already receiving pension. There was an empathetic comment on behalf of a nurse who "was trying to cope at work" even with diminishing power, reflecting the frailty-narrative (Taylor & Earl 2016), but they were also mentioned being "unpleasant … they should already retire" and also their places should already have been given to a younger worker.

Descriptions of the time on retirement, pensions and the pensioners

The time on old age retirement was described in the literature, on one hand, as a long awaited leisure where your own work has earned you pension (e.g. Ekerdt 2004, Smith & Dougherty 2012, Airio & Kangas 2017), and on the other hand, as the life phase with declining health (e.g. Karisto 2008). Among the retired writers on the forum, only the leisure-narrative was experienced. Retirement was mentioned as the ability to "fulfil the dreams" and have time to rest. Pension was a distinct part of the descriptions. It was told that the three things that are most important in life are "bank account, pension [and] extra-pension", because "when there is enough money [during retirement], the world is open". Those who were more moderate said, they sleep better when the loans are not distressing.

The experienced negative aspects of retirement were not that health was deteriorating, but that it was boring, and that one could not be useful. The disability pensioners that there were in the data were mainly concerned about the financial issues their inability to earn a salary caused them. A young disability pensioner commented that "as 28 years of age, to spend the rest of my life with but 700 euros' pension benefit does not appeal."

Other writers' gave pensioners many different, often contradictory, roles on the forum. They were considered vulnerable, and associated with less advantageous social groups such as the poor and the unemployed: One writer was worried about the national economics going bad "for the sake of the unemployed, the poor and the pensioners", and another writer claimed that those with small pension should have "more purchasing power". On the other hand, retirees' importance as providers for the others, as taxpayers, consumers and volunteers and as helping their adult children, were noticed. Not all were considered poor for there were also those having "too big pensions". Whether or not these different types of pensioners had retired successfully (Havighurst 1961) will be analysed in the next section.

5.3.3 The main narrative of retirement in the forum: *Retirement is the final line in the career race*

The main narrative of retirement found from the forum was that retirement was a matter of success and failure. The first trace of the narrative arose from a writing boasting about the writer's great success in life. They were "36 years old [and] on retirement", and unlike other early retired persons of the forum, this was not due to illness nor layoff. Instead, they had retired after they had earned a fortune through "working, business transaction and natural talent", and could now "live here on other side of the globe where there is warm", able to "pursue [their] dreams". It sounded like the person was proud they had succeeded in landing a more comfortable, earlier retirement than most other people, as if there were a game: use your years at paid work to maximise the length and quality of your retirement days (see e.g. Osborne & Rubinstein 1994).

This emerging narrative was called the narrative of *retirement as the final line in the career race* (later also *career-retirement game* or simply *retirement game*). It became even more appealing when it was understood being in line with the narrative of *successful retirement* found in previous studies (from Havighurst 1961 onwards).

The narrative of successful retirement in Finland, as described in the second chapter, has been characterised in the literature by a decent pension and plenty of healthy,

happy years on retirement, in pleasing activities (Kujala 2006, Frantsi 2012). Why the corresponding narrative found on the forum was not called that of successful retirement is because the frame of viewing it as the final line offers a better perspective to view retirement as a social process (e.g. Olds et al. 2018). Narrative of successful retirement concentrates in analysing living and life quality on retirement (e.g. Havighurst 1961). The narrative of career-retirement game that emerged from the forum reaches over the whole career where the strategies to achieve success are designed. In the narrative of successful retirement, career is also mentioned as a life phase necessary to provide whatever is needed in retirement. The difference is, on the forum, there were less descriptions of what is it like to be a successful pensioner, and more discussion on how this should be achieved. Also, the narrative found from the forum was expected to be different from that maintained in the wider context of the society. By giving the narrative a different name it was easier to trace its distinct features and minimise the confirmation bias, i.e. that the prejudices of the researcher would define what results are found.

In the context of this narrative of career-retirement game, successful retirement is called "victory in the game", working and planning for retirement as "playing the game" and dissatisfaction with retirement as "losing the game". These terms reflect only the attitudes writers on the forum shared; by no means are they the authors' views of how one should or should not retire.

Success in the game

After finding one conversation spreading the narrative of retirement as the final line in the career race, other discussions were eyed in search for traces of this narrative and *fractures,* or inconsistencies in it (Smith & Dougherty 2012). It was hypothesised, in the forum, success and failure in retirement game could be described in various ways, and none of them needed to be those of the established narrative of successful retirement discussed in Chapter 2. So, a theory of how to recognise a narrative of success or failure was logically constructed.

Success and failure in a game, in general, are defined by the rules of that game (Osborne & Rubinstein 1994). In the retirement game, the written rules are the legal regulation (macro-level) and workplace practices (meso-level). They indicate that you should work and pay pension payments until pension eligibility age. The other type of rules in the socially constructed normative game of retirement are the moral, or the oughtness norms (Hechter & Opp 2000), which are harder to define as such. What can be defined, though, is that following a norm leads to social acceptance. Thus, succeeding in following an oughtness norm should result in approving if not praising

comments. Likewise, disapproval and mockery would indicate a failure. (Xenitidou & Edmonds 2014.) So when reading and unravelling the narrative of success and failure in the career-retirement game, the gaze was focused onto parts of the conversations that expressed social (dis-)approval. The things that are defined as successful in the established narratives of successful retirement were ignored during this part of the analyses (unless they were socially approved in the forum), and only later when the narrative of retirement game was fully formed were they compared to the characteristics of the success/failure in the career-retirement game.

To retire early and rich was, as mentioned, the first type of winning the retirement game recognised from the forum, and for many writers, pension was a big thing as described in the section Transition to retirement. To these writers, the ultimate victory is thus expected to be theirs who can retire the earliest with the most money.

But when the conversations were thoughtfully read, there were also more modest ideas about the importance of pension, similar to Kujala's study's participant's, comment that "money is not so important to me" (2006, 133). One of the writers on the forum says that 1000 euros of pension would suffice a grandma, and another estimates that any pensioner should fare with 2000 euros a month. The things to do on retirement nor the things that make you happy are not necessarily related to pension size. Indeed, when a writer is asked for the three things in life that they are most thankful for, they claim that they "could list at least 30 things or more. And none of these is related to my pension being small or big". Other writers list free time, sleep and social relationships as the most important things in happiness. Second 'winner' that was found was a writer who described their retirement contently as a period in life during which they finally could "do whatever you want, sleep if you feel sleepy, do your own thing without work intervening". Money was not mentioned to be in a major role for them.

The third type of winners (or rather, will-be winners) on the forum were the writers, introduced earlier with the norm of retirement age, who were planning for retirement at pension eligibility age. They were, unquestioned, given the advice they needed, signalling social acceptance for they were following the norm (see also Kujala 2006).

Social acceptance was received in the forum also by keeping up in one's job, following what was the duty as in the example of the nurse who was working despite their pains. They had not yet made it to the final line, though, so it seemed they had not yet established their victory. Would they fail to keep up in their job, they, as many other writers, would end up into the disability retirement: one of the recognised ways to lose the career-retirement game.

Failure in the game

As there were winners in the game of how to retire, I looked for 'losers' as well; losers, as the emerging narrative of career-retirement game described them in the forum. It has to be noted that the main reason they are 'losers' is only because the discussion forum's (and probably also the society's) cultural narrative is disapproving their retirement. If there were a different master narrative in the society, they could be the winners in the game (Van Sluytman & Torres 2014). Alas, one writer thought that meaningful life ends whit the start of "the boring retirement days". For them, there could only be losers in the game.

This was not the general view in the forum, however, so losers had to be distinguished by other means than based on the fact they had reached the final line, i.e. had retired. Loosing, in this context, is defined by the inability to follow either formal or moral norms of the society (Hechter & Opp 2000), which can be recognised from inability to reach pension eligibility age with enough pension (the formal norm) or from the self- or social disapproval (the moral norms). Thus, a loser would be someone who did not boast about their success in retirement but were pitying themselves, or someone who was disapproved by other writers' comments.

The simplest way of losing the retirement game was, inversely to the *American dream* -narrative of successful retirement, to not be able to accumulate big enough pension by the time one retired. This was the destiny of a writer who, due to their birth cohort, had fallen out of the accumulation of pension during child care and work as a young person, and would not receive the increased accrual rate that was given to some cohorts during the pension reform 2005.

Deteriorating health was mentioned in four conversations as the reason why retirement was about to happen or had already happened without the person wanting it. The most extreme case was a very young person who was "trying to get on in my life... At the age of 28, it does not seem too tempting to spend rest of my life with a [disability] pension of 700 euros... But what could be my profession?" Due to their illness, they had been forced to discontinue their studies in university, and looking at the job ads made them scared that there was no one profession they could succeed in. Thus, the game was lost even before it had really started.

The phenomenon that "so many young people have to retire due to disabilities" was considered "miserable" by another writer. What is meant by "young" is not clear, but the 28-years-old will definitely fall into that category. More common in the forum was that disability retirement was threatening those already in working life. One writer was proud that they had "been working nearly all my life". But then, at the age of 58, they had "had to retire due to illness". A nurse was told to "hang on [in her job] towards

retirement age" in order to avoid disability retirement. It was not evident if the reason for disregarding one's own disability retirement was the failure to reach the normative retirement age or if it was the smaller pension benefit that it meant. It was clear, however, that disability retirement was something people on the forum were trying to avoid. Despite it is not rare to retire through disability retirement (ETK 2017), this norm does not provide its followers with a pleasant role.

Another involuntary reason for retirement was layoff. Two writers recognised it is a phenomenon that older workers face age discrimination from their employers: "Many would be ready to continue their careers until 70 years of age but the employers won't let them", "older workforce is compelled into forced retirement". Fear of being forced on to early retirement had made one writer to explore the possibilities how it would affect her pension: "[my work place] is under cooperation negotiations, age almost 62, personal retirement age 64. If I [follow different scenarios], will I lose anything [of my pension] ..." The latter writer is directly mentioning the 'loss' that they will face due to forced early retirement.

Besides these failures to meet the formal norm of retirement, there were moral losers of the game, too. Pensioners who worked were considered guilty of taking jobs from the young, thus were considered breaking the moral norms of retirement. Also, central part of the narratives of successful retirement, both in previous literature and on the forum is, pension has to be earned through working (Airio & Kangas 2017, Smith & Dougherty 2012). A retired farmer was strongly disapproved by a writer for the writer thought, the farmer had never really worked nor paid pension payments and yet, had retired and received a large pension.

Playing the game: the sub-narrative of *Respectful persistence at work until pension*

Before a victory or a loss in the retirement game, there is the part that affects the result: the playing. Playing in any game means making strategic choices guided by the rules in order to achieve the wanted end result (Osborne & Rubinstein 1994). In the retirement game, the rules come from the norms of career and retirement and they set the limits for playing strategies. Examples of how the written rules enact were seen in the forum: the 'retirement wizard' had to answer questions about them, some writers were frustrated over them when they treated them unfairly, and hopes they would be changed were present in three discussions, namely that small pensions should get more purchasing power, and, contradictory, the pension eligibility age should be lower, but also older workers should not be forced into retirement if they still could continue. Also the meso-level rules were visible in the form of layoffs of older workers. Mainly the

rules were taken as granted and obeyed, but, when there was chance for flexibility, individuals on the micro-level started playing by choosing different strategies.

In the forum, the most popular strategy for achieving success on the final line in the race, e.g. decent pension with minimal years at work and with social acceptance, was to work for salary until but no longer than the pension eligibility age. This formed the sub-narrative of *respectful persistence at work until pension*: Retirement was obvious but so was work until that (see also Kujala 2006). The best examples of this were the writers introduced in the section Retirement choices and the norm of retirement age who wished for a pleasant job [just] until pension eligibility age and who, when it seemed they could not keep in their current work until they would achieve retirement, began to plan for changing career into one that could be done until pension eligibility age. Even when the work caused pain, as it did for the nurse presented earlier, their survival was respected. When another person retired earlier than what their retirement age were, the event was described as "[they] had to move to disability retirement". To not work had to be explained with some socially accepted reason; and usually this was disability or lay-off.

Work after pension eligibility age was strictly against the moral rules of the game, which could be notices by the strong criticism this behaviour received (see section on moralising other's choices). Working pensioners were found curious if not malevolent because they were blocking the access of the young people to (scarce) jobs. From the perspective of the race for retirement, this equals to hacking the results of the game after the final line, which means it is possible to keep playing even after retirement.

On the opposite side of the working pensioners, there were writers on the forum who did not share and even resisted the narrative that all should work until their proper retirement age. The rich writer who was enjoying their retirement in somewhere warm thought it proper enough to gain a fortune and retire, despite age. Those older workers who were afraid of layoffs did not even think of seeking another job if they would have to leave their current job. These notions created a fracture within the narrative (Smith & Dougherty 2012), and the fracture was enlarged by those who viewed the older workers as blocking the young. Also the writer commenting on the poor customer service they received thought the workers should not wait for their retirement any longer. These fractures show there are other cultural narratives and norms competing with the narrative of respectful persistence: that the rich have different rules than others in the society; that after a certain age no-one will hire you; that older people are not as valuable as the young; and that older workers are less competent than the young.

Other playing strategies were also present in the forum. The entrepreneurs were free or almost free to choose the timing of their retirement regardless of the pension eligibility age because of other sources of income they possessed. If the comment on the farmer is factual, it was also possible to cheat in the game by "letting tax-payers pay for their pensions". Also, the third entrepreneur who was mentioned in the forum was not even playing the game as they were not paying enough pension payments to add up to decent pension benefit. Not needing to play was overall considered a positive outcome of the game, but the opposite also held: Inability to play the career-retirement game was the destiny of those with failing health or a layoff. Their situation, when discussed on the forum, was always found unfortunate.

Representing contradictory "truths": The side narrative of *National economy's future depends on retirement age*

A small, independent narrative of the linkage between national economy and retirement was also found from the forum. National economics and retirement age were seen interconnected but their relation was disputed. On one hand, the too high statutory retirement age was associated with a high rate of unemployment, for which reason "[r]etirement age should not be lifted". Using pension funds instead of unemployment benefits was also seen a more humane way of controlling older adults' transition out of working life: "Retirement age had to be lifted when it was claimed that pension funds will not suffice, but bearing in mind that the baby boomers already retired, the increasing pension funds are telling a wholly different story. Now those who would already have deserved their pension are clinging on to working life because otherwise they would end up to [unemployment] and to smaller pensions". The latter writer understood that pension funds are separate from the tax-based unemployment benefits and suggested, if the 100 000 workers and unemployed over 60 years old were moved to retirement, "tens of thousands work places were freed and the state would save unemployment benefits". On the other hand, one writer wondered, "is it not on everyone's lips that the retirement age should be increased otherwise the national economy will not be rebuilt."

So there were two contradictory 'truths': either retirement age should not be raised and this would save the country from older workers' unemployment and from the costs related to that, or retirement age should be increased so that there would be a larger labour supply to provide for the nation. Both claim that this would solve the issues there are in the national economy. The two contradicting narratives of the effect that retirement age has on the national economics cannot both be true. Instead, they are a neat summary of the trustworthiness of the narratives spreading on the forum.

Nevertheless, both of them are seen by possibly thousands of visitors who, unless they can resist the narrative, may easily adopt the views offered in them as truths.

5.4 Reflections on the qualitative analyses

The data for the qualitative analysis was selected using the results of a topic model. This kind of a mixed methods approach for analysing large quantities of text is not yet mainstream in the Social Sciences. The typical approach would be taking a sample, which is less time consuming than topic modeling the data. By comparing the results of the pre-study and the later analyses it will be analysed, was topic modeling a better tool for finding relevant content among the conversations. The method for choosing the conversations for the qualitative analyses will also be briefly discussed.

5.4.1 Topic modeling in finding data: Pointless effort or a necessary tool?

In the pre-analysis of the forum in third chapter, the random sample of 11 conversations from the Finland24 forum, together with the codes and themes found from there, was introduced. There were multiple reasons why a sample was analysed before constructing a topic model. It guided the analytical choices in cleaning the data, valuated at least partially that mention of the word 'retirement' or 'pension' would yield discussions related to those two, and, most importantly, provided material for analysing was the topic modeling of the data even necessary. Ylisiurua (2017) noted, it would have been possible to take a sample of the messages in the forum and analyse them instead of structuring the data by topic modeling. She considered this impractical as the data is so vast that close-reading as the only method could possibly produce one interesting perspective to the data but not capture the general view. (Ibid., 59.) She did not, however, perform such close-reading but only hypothesized the result which is why it remained untested until now: when the data is vast, can topic modeling find such information among the data that would be missed by taking a random sample?

The aim of topic modeling the data was to limit the analyses to the conversations that would include traces of retirement narratives. From all the topics the topic model output it was considered, those related to the theme of *Retirement transition* would best suit as the base for the analyses as retirement narratives incarnate in the norms of how the transition should be done properly (see Chapter 2).

The conversations in the two topics that were analysed as being about retirement transition were, when investigated by the author, truly about the issue: 9 out of 28 had retirement transition as the major issue in the conversation and 4 more mentioned it. Compared with the discussions presented in the Chapter 3, there was one discussion there that discussed the transition to disability retirement, and another suggested that pensioners should be taken back to work, but none were about the transition to old age retirement.

Thus, the model fared better in finding relevant discussions for the qualitative analysis than what a random sample did, as Ylisiurua (2017) had estimated. Yet, using topic modeling as a method for sorting the data means that there are two data and two analyses in the study: first cleaning and topic modeling the whole of the data, labelling and grouping the topics and selecting the (possibly few) relevant topics, then using the documents within the relevant topics as the data for the qualitative analysis. Thus, topic modeling is not a fast and easy method for sorting the data, but if the purpose of the study is to find specific information out of massive data, it still is better than taking a sample - not to mention how drastically faster it is than human reading and sorting that size of a data (see e.g. Quinn et al. 2010).

5.4.2 How to choose the conversations after topic modeling

After topic modeling is ready, how should the conversations be selected for the analysis? Scholars combining topic modeling and qualitative analysis have used different methods for this, from keyword searches (e.g. Ylisiurua 2017) to using all of the conversations within the selected topics (e.g. Ylä-Anttila 2018). In this study, only 28 conversations were used in analysing the retirement norms and narratives of the forum. It was hypothesised, based on the results of topic modeling, and selecting carefully the most representative conversations from the selected topics, would yield the analysis the most relevant of all of the forum.

On one hand, the qualitative analyses based on the selected conversations were successful in that retirement narratives could be found from the conversations, and in that the found narratives were consistent with, yet provided existing knowledge with new information about retirement narratives in Finland. On the other hand, there were large parts in the conversations that needed to be coded as irrelevant to the study objective, and themes that were not directly about retirement transition were also found from the conversations. Thus, the careful selection of the messages provided the analysis with mainly relevant material while some parts of it were not useful.

The question of how to choose the conversations for further analyses is one of the questions that needs to be solved in order to increase the replicability of the studies utilising topic modeling for selecting data. However, all the existing methods used for retrieving the documents have provided relevant results. This was also the case with this study: retirement narratives could be found from the selected conversations. As long as the method for selecting the documents is clearly stated, replicating the study should be possible. There might still be a universally best method for this task, which is why further research on this area is necessary.

6 Conclusions and further research objectives

The research questions set for this study were two-sided. On the one hand, there was a genuine interest into the retirement discussion on Finland's largest online discussion forum. On the other hand, the online forum with its massive amount of conversations offered a chance to present and test the novel method of computational topic modeling.

In this conclusive section, the two-folded results of this Master's thesis will be summarised. First, the normative narratives found from Finland24 are presented. The analysis of the narratives was limited on only a small section of the vastness of the discussion forum, yet the results signalled that at least some narratives were recognisable. However, further research needs remain. Nor is the establishment of topic modeling as a method in Social Sciences ready, and these research gaps will be discussed in the second part of this chapter.

6.1 Finland24 and retirement narratives

Norms and narratives related to retirement guide individuals' retirement expectations and decisions (Xenitidou & Edmonds 2014). Traditionally, these have been studied by asking people to tell about their retirement intentions and analysing the resulting texts and or interview recordings (e.g. Kujala 2006, Frantsi 2012). In this study, a different approach was taken: what if the narratives could be found from an online discussion forum? Would these narratives be the same as the ones received by the traditional methods, or reflect unnoticed perspectives?

These questions were answered in Chapter 5. Truly, the conversations in the forum intermediated norms and narratives of retirement. The argumentation there was typically based on "common knowledge", that is, culture-wide narratives of the society, and the writers valued other people's retirement and non-retirement decisions based on their normative expectations. An especially strong norm there was the norm of pension eligibility age as the right timing for retirement. Paid work after that was generally disapproved. The norms that governed how time spent on retirement and pensioners should be like were more flexible, allowing different terms of the act of retirement to be used and contradictory views of retirement as awaited leisure or as unwanted and boring.

The main narrative of retirement that was found from the forum was close to that of successful retirement found in previous studies (e.g. Havighurst 1961), where success in retirement depends on the cultural context of the narrative but is typically defined by

life satisfaction on retirement. In the forum, however, the narrative reached more holistically not only over retirement but also on the whole career, as the process-nature of retirement would imply (Olds et al. 2018). The career-part formed its own sub-narrative within the main narrative of career-retirement game. In this narrative, career was seen as the time when the right choices have to be made so that the norm of retirement at pension eligibility age is reached. Successful retirement days meant both big, or at least enough pension, and freedom to choose over how to spend one's time.

What separated the 'winners' of the career-retirement game from the 'losers' of it was social- and self-acceptance of the trajectory that lead to retirement. If there had been disruptions in the career, that meant smaller pension accumulation. Economic difficulties during retirement were the main reason for not finding oneself successfully retired. Another way of losing was through failing the social respect and thus the normative game of retirement by cheating in the game, which was done by making others pay for one's pension.

Why the main narrative of retirement found from the forum was more concentrated on the whole career than those found in previous literature? This can only be speculated. One reason could be that there has yet been too little research on retirement narratives in general, and all the different aspects of them have not yet been fully reached. How central part career is in successful retirement can only be solved with further research, possibly so that traditional interview-based methods are combined with analyses of self-producing social media data. Another hypothesis is, the narratives of retirement that are told in face to face interaction or even in questionnaires are different to those described from behind the shield of online-anonymity and without anyone asking. But which of them are more real, and which affect the individuals in the society more? This question will also be left open for futher research to answer. In any case, both of them are part of the reality, and their norms affect people living in the middle of these narratives. Because it is unclear if the main narrative found from the forum really is a culture-wide narrative, it is not called a master narrative.

Overall, the analyses of the norms and narratives in the forum concentrated on analysing but a small part of all the conversations (28 of them) that there are about retirement on the forum. Further research would be needed to find to what extent the career-retirement game would hold over all of the forum. Also, other retirement narratives might thus appear.

6.2 Topic modeling and Social Sciences

Overall, it was noticed, the studies using topic modeling have this far mainly focused on developing and analysing the method, or they have simply analysed the topics it outputs without a deeper look into the qualitative aspects it could reveal from the data. So this study would also show how topic modeling can be used as a pre-processing step to access the interesting themes in a corpus, which may then be analysed qualitatively.

In this thesis, special effort was put in developing a more systematic approach to labelling the results of the topic model that any other study has so far done. The need for this is because in that part of topic modeling the risk for a subjective interpretation is the greatest. By describing in detail the reasons why specific topics were labelled, and why others were left out of the further analysis the process becomes more transparent and is easier to replicate. Also the other analytical choices made in this study have been tried to describe in detail for the same reasons.

Before the thesis, the benefits of topic modeling as a data selection method had not been compared with those of taking a random sample. The qualitative pre-analysis enabled a modest comparation, where topic modeling seemed to provide the later qualitative analysis with more relevant data than what a random sample had done. This analysis had its limitations, and it should be repeated with different data and bigger sample before anything conclusive could be said about its superiority/inferiority.

All the analysis together could not give a complete picture of what is the retirement discussion in the forum, and how does it reflect the society-wide ideas people have of retirement. Only a small fraction of the results of the topic modeling were analysed. Besides the theme Retirement transition that was selected into the qualitative analyses, there were four distinctly retirement-related themes found. How these aspects of retirement are discussed in the forum could offer relevant information for the designers of the retirement system as well as other scholars studying retirement. Overall, the treasures the Finland24-data could give for Social Sciences should be further investigated, and the type of mixed methods approach that was used in this thesis has proven to be a useful tool for this task.

References

Academy of Finland (2017) Citizen Mindscapes – Detecting Social, Emotional and National Dynamics in Social Media. [Project brochure.] Retrieved January 12, 2017, from http://www.aka.fi/digihum.

Achenbaum, W. Andrew (2006) What is retirement for? The Wilson Quarterly, 30(2), 50–56.

Adams, Gary, Beehr, Terry (eds.) (2003) Retirement: Reasons, processes and results. New York: Springer Pub.

Ahonen, Pertti (2015) Institutionalizing Big Data methods in social and political research. Big Data & Society, July-December 2015, 1–12.

Airio, Ilpo ja Kangas, Olli (2017) Mielipiteet eläkkeiden järjestämistavoista ja määräytymisperiaatteista Suomessa vuosina 1994 ja 2016. Yhteiskuntapolitiikka 82(6), 617–629.

Aller Media Oy (2014) Suomi 24 -korpus (2016H2). [Text corpus]. Kielipankki. Available: http://urn.fi/urn:nbn:fi:lb-2017021506

Aller Media Oy (2016) Suomi24.fi: Suomen suurin verkkoyhteisö. [Project brochure.] TNS Metrix on Google Analytics, weeks 41–52/2015. Information received by e-mail.

Auerbach, Carl & Silverstein, Louise (2003) Qualitative data: An introduction to coding and analysis. New York, NY: New York University Press.

Baumer, Eric P., Mimno, David, Guha, Shion, Quan, Emily, & Gay, Geri K. (2017) Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? Journal of the Association for Information Science and Technology, 68(6), 1397–1410.

Berger, Peter L., Luckmann, Thomas (1991) The social construction of reality: A treatise in the sociology of knowledge. London: Penguin UK. (Original work published in 1966.)

Blei, David M., Andrew Y. Ng, Michael I. Jordan (2003) Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993–1022.

Blei, David M., McAuliffe, Jon D. (2007) Supervised topic models. [A seminar paper.] The twenty-first Annual Conference on Neural Information Processing Systems. Hyatt Regency Vancouver, Vancouver, B.C., Canada.

Blei, David M. (2012) Probabilistic topic models. Communications of the ACM. 55(4), 77–84.

von Bonsdorff, Monica, Huuhtanen, Pekka, Tuomi, Kaija and Seitsamo, Jorma (2010) Predictors of employees' early retirement intentions. An 11-year longitudinal study. Occupational Medicine 60(2), 94–100.

Brauer, René, Dymitrow, Mirek (2014) Quality of life in rural areas: A topic for the Rural Development policy? Bulletin of Geography. Socio-economic Series, 25(25), 25–54.

Chang, Jonathan, Gerrish, Sean, Wang, Chong, Boyd-Graber, Jordan L., Blei, David M. (2009) Reading tea leaves: How humans interpret topic models. Advances in neural information processing systems 22 (NIPS 2009), 288–296.

Chuang, Jason, Wilkerson, John D., Weiss, Rebecca, Tingley, Dustin, Stewart, Brandon M., Roberts, Margaret E., Poursabzi-Sangdeh, Forough, Grimmer, Justin, Findlater, Leah, Boyd-Graber, Jordan, Heer, Jeffrey (2014) Computer-assisted content analysis: Topic models for exploring multiple subjective interpretations. Advances in Neural Information Processing Systems Workshop on Human-Propelled Machine Learning.

DiMaggio, Paul, Nag, Manish, & Blei, David (2013) Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. Poetics, 41(6), 570–606.

Doig, Christine [PyTexas] (2015, October 15) Topic Modeling with Python. Retrieved from https://www.youtube.com/watch?v=BuMu-bdoVrU.

Ekerdt, David (2004) Born to Retire: The Foreshortened Life Course. The Gerontologist, 44(1), 3–9.

Ekerdt, David (2010) Frontiers of research on work and retirement. Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 65(1), 69–80.

Elovaino, Marko, Forma, Pauli, Kivimäki, Mika, Sinervo, Timo, Sutinen, Risto, Laine, Marjukka (2005) Job demands and job control as correlates of early retirement thoughts in Finnish social and health care employees. Work & Stress, 19(1), 84–92.

Eskola, Jari, Suoranta, Juha (1998) Johdatus laadulliseen tutkimukseen. Vastapaino.

Esping-Andersen, Gøsta (1990) The three worlds of welfare capitalism. Princeton, NJ: Princeton University Press.

ETK (2017) Tilasto Suomen eläkkeensaajista 2016. Eläketurvakeskuksen tilastoja 11/2017. Helsinki: Juvenes Print - Suomen Yliopistopaino Oy.

ETK (2018) Eläketurvakeskuksen tutkimuksia -blog. Available in https://www.etk.fi/blogit/pdf\s\do5(c)at/tutkimuksia/. (Read 18.3.2018)

Frantsi, Tapani (2012) Ikääntyvän johtajan kilvoittelevat valinnat. Gerontologia, (26)2, 86–99.

Furunes, Trude, Mykletn, Reidar J., Solem, Per Erik, de Lange, Annet H., Syse, Astri, Schaufeli, Wilmar B., Ilmarinen, Juhani (2015) Late Career Decision-Making: A Qualitative Panel Study. Work, Aging and Retirement, 1(3), 284–295.

Gilleard, Chris (2007) Old Age in Ancient Greece: Narratives of desire, narratives of disgust. Journal of Aging Studies, 21(1), 81–92.

Grimmer, Justin, Stewart, Brandon M. (2013) Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political Analysis, 21(3), 267–297.

Griffiths, Thomas L., & Steyvers, Mark (2004) Finding scientific topics. Proceedings of the National Academy of Sciences, 101(Supplement 1), 5228–5235.

Gullette, Margaret Morganroth (1997) Declining to Decline: Cultural Combat and the Politics of the Midlife (Age Studies). Charlottesville, VA: University of Virginia Press.

Hakala, Salli, Vesa, Juho (2013) Verkkokeskustelut ja sisällön erittely in Laaksonen, Salla-Maara, Matikainen, Janne, Tikka, Minttu (eds.), Otteita Verkosta. Verkon ja sosiaalisen median tutkimusmenetelmät (216–244). Tampere: Vastapaino.

Havighurst, Robert (1961) Successful Aging. The Gerontologist, 1(1), 8–13.

Hechter, Michael, Opp, Karl-Dieter (2001) What have we learned about the emergence of social norms? In Opp, Karl-Dieter, Hechter, Michael (eds.), Social Norms (394–416). New York, NY: Russel Sage Foundation.

Järnefelt, Noora (ed.) (2017) Työolot ja työurat: tutkimuksia työurien vakaudesta ja eläkkeelle siirtymisestä. Eläketurvakeskuksen tutkimuksia 08/2016.

Järvensivu, Anu & von Bonsdorff, Monika E. (2018) Yrittäjien eläköitymisprosessien moninaisuus – siirtymisprosessin dynaamisuus ja operationalisointi. Työpoliittinen aikakauskirja 4/2018, 35–47.

Kallio, Sanna (2013) Antaa työn imun viedä eläkeikään asti. [Master's thesis.] University of Helsinki, Helsinki, Finland.

Karisto, Antti, Konttinen, Riikka (2004) Kotiruokaa, kotikatua, kaukomatkailua. Tutkimus ikääntyvien elämäntyyleistä. Helsinki: Palmenia-kustannus.

Karisto, Antti (2007) Finnish Baby Boomers and the Emergence of the Third Age. International Journal of Ageing and Later Life, 2(2), 91–108.

Karisto, Antti (2008) Satumaa. Suomalaiseläkeläiset Espanjan Aurinkorannikolla. Helsinki: Suomalaisen Kirjallisuuden Seura.

Koll, Matthew B. (1979) WEIRD: An approach to concept-based information retrieval. In ACM Sigir Forum, 13(4), 32–50.

Komp, Kathrin, Aartsen, Marja (eds.) (2013) Old Age in Europe. A Textbook of Gerontology. Dortrecht: Springer Netherland.

Koskela, Annukka (2019, April 6) Eläköityminen lisää naisilla alkoholin käyttökertoja, miehillä annosten määrää. Helsingin Sanomat, Retrieved from www.hs.fi.

Kujala, Tiina (2006) ”Ei pirise enää koulun kello”: kerronnallinen tutkimus opettajien ikääntymiskokemuksista. [Doctoral dissertation.] Tampere: Tampere University Press.

Kyyrä, Tomi, Paukkeri, Tuuli (2015) Työkyvyttömyyseläkkeiden maksuluokkamallin kannustinvaikutukset. Eläketurvakeskuksen raportteja 07/2015.

Laaksonen, Salla-Maaria, Matikainen, Janne, Tikka, Minttu eds. (2013) Otteita verkosta. Verkon ja sosiaalisen median tutkimusmenetelmät. Tampere: Vastapaino.

Lagus, Krista, Pantzar, Mika, Ruckenstein, Minna, Ylisiurua, Marjoriikka (2016) Suomi24. Muodonantoa aineistolle. Kuluttajatutkimuskeskus 2016:10. Helsinki: Valtiotieteellisen tiedekunnan julkaisuja, Helsingin yliopisto.

Lahti, Vesa-Matti, Hämäläinen, Mirja (2017, June 30) Martha Nussbaum ja Saul Levmore haluavat ikääntyä viisaasti. [Blog post] https://www.sitra.fi/artikkelit/martha-nussbaum-ja-saul-levmore-haluavat-ikaantya-viisaasti/ .

Lau, Jey Han, Newman, David, Baldwin, Timothy (2014) Machine Reading Tea Leaves: Automatically Evaluating Topic Coherence and Topic Model Quality. Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, 530–539, Gothenburg, Sweden.

LaValley, Barry & Finke, Mark (2014) So You Think You Are Ready To Retire? What You REALLY Want to Know Before You Take The Leap! Retirement Lifestyle Center Inc.

Levy, Karen E. C., Franklin, Michael (2014) Driving Regulation: Using Topic Models to Examine Political Contention in the U.S. Trucking Industry. Social Science Computer Review, 32(2), 182–194.

Leydesdorff, Loet, Hellsten, Iina (2006) Measuring the meaning of words in contexts: An automated analysis of controversies about 'Monarch butterflies', 'Frankenfoods', and 'stem cells'. Scientometrics, 67(2), 231–258.

Lock, Andy, Strong, Tom (2010) Social Constructionism. Sources and Stirrings in Theory and Practice. Cambridge: Cambridge University Press.

Lucas, Christopher, Nielsen, Richard A., Roberts, Margaret E., Stewart, Brandon M., Storer, Alex, Tingley, Dustin (2015) Computer-assisted text analysis for comparative politics. Political Analysis, 23(2), 254–277.

Lyotard, Jean-François (1984) The Postmodern Condition: A Report on Knowledge. Theory and History of Literature, 10. Manchester: Manchester University Press. (Original work published 1979.)

McCallum, Andrew Kachites (2002) MALLET: A Machine Learning for Language Toolkit. http://mallet.cs.umass.edu.

Merlino, Kathy (2018, September 3) Exactly What Is Retired? [Blog post] Retrieved from https://kathysretirementblog.com.

Mimno, David, Wallach, Hanna M., Talley, Edmund, Leenders, Miriam, & McCallum, Andrew (2011) Optimizing semantic coherence in topic models. In Proceedings of the conference on empirical methods in natural language processing (262–272). Association for Computational Linguistics.

Moen, Phyllis (2012) Retirement dilemmas and decisions. In Hedge, Jerry W. and Borman, Walter C. (eds.) The Oxford Handbook of Work and Aging. Oxford: Oxford University Press.

Mohr, John, Bogdanov, Petko (2013) Topic models: What they are and why they matter. Poetics, 41(6), 2–40.

Mookiah, Lenin, Eberle, Willliam, Siraj, Ambareen (2015, April) Survey of Crime Analysis and Prediction. In FLAIRS Conference, 440–443.

Nelimarkka, Matti (2019) Aihemallinnus sekä muut ohjaamattomat koneoppimismenetelmät yhteiskuntatieteellisessä tutkimuksessa. Politiikka: Valtiotieteellisen yhdistyksen julkaisu, 61(1), 6–33.

Niemi, Matti (2016) Kansainvälinen verotus ja globaalin talouden reterritorialisaatio. Aihemallinnus ei-valtiollisten toimijoiden lausunnoista OECD:n BEPS-hankkeessa. (Master's thesis.) University of Helsinki, Helsinki, Finland.

Nivalainen, Satu, Järnefelt, Noora (2017) Eläkeaikeet ja toteutunut eläkkeelle siirtyminen: sukupuolen, työskentelysektorin ja muuttuneiden elämäntilanteiden vaikutukset. Eläketurvakeskuksen tutkimuksia 06/2017.

Nuortimo, Jussi (2016) Of spaceships and men - Mapping the polity of EVE Online using Latent Dirichlet Allocation topic modeling. [Master's thesis.] University of Helsinki, Helsinki, Finland.

Oinas, Tomi, Anttila, Timo, Nätti, Jouko (2016) Parempi työn laatu, pidemmät työurat? Työn laadun vaikutus varhaiseen eläkkeelle siirtymiseen Suomessa. In Järnefelt, Noora (Ed.), Työolot ja työurat (127–145). Tutkimuksia työurien vakaudesta ja eläkkeelle siirtymisestä. Eläketurvakeskuksen tutkimuksia 2016:8.

Olds, Tim Burton, Nicola W., Sprod, Judy, Maher, Carol, Ferrar, Katia, Brown, Wendy J., van Uffelen, Jannique, Dumuid, Dorothea (2018) One day you'll wake up and won't have to go to work: The impact of changes in time use on mental health following retirement. PLoS ONE, 13(6), 1–14.

O'Nan, Stewart (2019) Henry, Himself. New York, NY: Viking.

Osborne, Martin J., Rubinstein, Ariel (1994) A course in game theory. Cambridge, MA: MIT press.

Pang, Bo, Lee, Lillian (2008) Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1–2), 1–135.

Piensoho, Tuula (2014, February 2) Eläkevalmennusta tarvitaan - tässä valmentaja! [Blog post] https://www.piensohotraining.fi/blogi/?k=el%C3%A4kevalmennus.

Plisson, Joël, Lavrac, Nada, & Mladenić, Dunja (2004) A rule based approach to word lemmatization. Proceedings of IS-2004, 83–86.

Purhonen, Semi, Toikka, Arho (2016) "Big datan" haaste ja uudet laskennalliset tekstiaineistojen analyysimenetelmät : esimerkkitapauksena aihemallianalyysi tasavallan presidenttien uudenvuodenpuheista 1935–2015. Sosiologia, 53(1), 6–27.

Quinn, Kevin M., Monroe, Burt L., Colaresi, Michael, Crespin, Michael H., Radev, Dragomir R. (2010) How to Analyze Political Attentien with Minimal Assumptions and Costs. American Journal of Political Science, 54(1), 209–228.

Rautava, Mari (2015) "Unelmani on kuolla saappaat jalassa." Fenomenografinen analyysi eläkeläisten ikääntymiseen liittyvistä käsityksistä. (Master's thesis). University of Tampere, Tampere, Finland.

Rawls, John (1971) A theory of justice. Cambridge, MA: Harvard University Press.

Reday-Mulvey, Genevieve (2000) Gradual Retirement in Europe. Co- published simultaneously in Journal of Aging & Social Policy, 11(2), 49–60; and in Francis G. Caro, Morris, Robert, Norton, Jill R. (eds.), Advancing Aging Policy as the 21st Century Begins. Philadelphia, PA: The Haworth Press.

Rehurek, Radim, Sojka, Peter (2010) Software framework for topic modelling with large corpora. In LREC Workshop on New Challenges for NLP Frameworks, 45–50.

Riekhoff, Aart-Jan and Järnefelt, Noora (2018) Retirement trajectories and income redistribution through the pension system in Finland. Social Forces, 97(1), 27–54.

Röder, Michael, Both, Andreas, Hinneburg, Alexander (2015) Exploring the Space of Topic Coherence Measures. WSDM'15, February 2–6, Shanghai, China.

Sarvimäki, Anneli (2013) Vanheneminen eri kulttuureissa ja etnisissä ryhmissä. In Heikkinen, Eino, Jyrkämä, Jyrki, Rantanen, Taina (eds.), Gerontologia (85–94). Helsinki: Duodecim.

Smith, Frances L., Dougherty, Debbie S. (2012) Revealing a master narrative: Discourses of retirement throughout the working life cycle. Management Communication Quarterly, 26(3), 453–478.

Solem, Per Erik, Syse, Astri, Furunes, Trude, Mykletun, Reidar J., De Lange, Annet, Schaufeli, Wilmar and Ilmarinen, Juhani (2016) To leave or not to leave: retirement intentions and retirement behaviour. Ageing and Society, 36(2), 259-281.

Sorsa, Tiia (2014) Sukupolvien välinen oikeudenmukaisuus ikääntyvän Suomen työeläkejärjestelmässä. (Unpublished Bachelor's thesis.) University of Helsinki, Helsinki, Finland.

Stevens, Keith, Kegelmeyer, Philip, Andrzejewski, David, Buttler, David (2012) Exploring Topic Coherence over many models and many topics. Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 952–961.

The Suomi 24 Corpus (2016H2) (n.d.) Retrieved from: http://metashare.csc.fi/repository/browse/the-suomi-24-corpus-2016h2/eb323320f44d11e6b70e005056be118e30dc4e74e4654a4a8b3e8789ef31c0d0/ .

Suomi24 Kumppanuus (n.d.) Retrieved from https://www.aller.fi/media-2/kumppanuuskonseptit/.

Takala, Esa-Pekka, Seitsamo, Jorma, Hensiksson, Lea, Härmä, Mikko, Gould, Raija, Nivalainen, Satu, Kuivalainen, Susan (2015) Ikääntyvien työhön osallistuminen. Kirjallisuusselvitys 2000-luvulla julkaistusta suomalaisesta tutkimuksesta. Tampere: Juvenes Print.

Taylor, Philip, Earl, Catherine (2016) The Social Construction of Retirement and Evolving Policy Discourse of Working Longer. Journal of Social Policy, 45(2), 251–268.

Tikanmäki, Heikki, Appelqvist, Jukka, Reipas, Kaarlo, Sankala, Mikko, Sihvonen, Hannu (2016) Lakisääteiset eläkkeet: pitkän aikavälin laskelmat 2016. Eläketurvakeskuksen raportteja 08/2016.

Tikanmäki, Heikki, Lappo, Sampo, Merilä, Ville, Nopola, Tuija, Reipas, Kaarlo, Sankala, Mikko (2019) Lakisääteiset eläkkeet: pitkän aikavälin laskelmat 2019. Eläketurvakeskuksen raportteja 02/2019.

Trethewey, Angela (2001) Reproducing and Resisting the Master Narrative of Decline. Midlife Professional Women's Experiences of Aging. Management Communication Quarterly, 15(2), 183–226.

Tuominen, Eila, Gould, Raija (2010) Tutkimustietoa lainmuutosten vaikutuksista ja muista työssä jatkamista selittävistä tekijöistä. In Uusitalo, Hannu, Kautto, Mikko, Lindell, Christina (eds.) Myöhemmin eläkkeelle. Selvityksiä ja laskelmia. Eläketurvakeskuksen selvityksiä 2010:1 (28–38). Helsinki: Edita Prima Oy

Työeläke (2005) Eläkettä alkoi karttua myös vanhempainvapaalta. News in https://www.tyoelake.fi/fi/nainelakemuodostuu/mistaelakekertyy/elakettalapsenkotihoidosta/Sivut/default.aspx (Cited 26.4.2017).

Valtion virkamieslaki (750/1994) 35§: www.finlex.fi.

Van Sluytman, Laurens G., Torres, Denise (2014) Hidden or uninvited? A content analysis of elder LGBT of color literature in gerontology. Journal of gerontological social work, 57(2-4), 130–160.

Wang, Mo (ed.) (2012) The Oxford Handbook of Retirement. Oxford: Oxford University Press.

Why are machine learning models called black boxes? (2017, August 17.) [Discussion forum thread.] Retrieved from: https://datascience.stackexchange.com/questions/22335/why-are-machine-learning-models-called-black-boxes

Wilson, Andrew T., & Chew, Peter A. (2010) Term weighting schemes for latent dirichlet allocation. In human language technologies: The 2010 annual conference of the North American Chapter of the Association for Computational Linguistics (465–473). Association for Computational Linguistics.

Xenitidou, Maria, Edmonds, Bruce (eds.) (2014) The Complexity of Social Norms. Berlin: Springer.

Yatsko, Adrzej & Suslow, Walery (2016) Insight into Theoretical and Applied Informatics. Introduction to Information Technologies and Computer Science. Berlin: Sciendo.

YLE (2017, May 9) Eläköityminen vähentää stressiä - mutta lähinnä vain hyväpalkkaisten. YLE. Retrieved from www.yle.fi.

Ylisiurua, Marjoriikka (2017) Aihemallinnuksen mahdollisuudet sosiaalisen median aineistojen jäsentämisessä – terveyskeskustelu Suomi24-verkkopalstalla. Kulutustustkimus. Nyt, 11(2), 44–67.

Ylä-Anttila, Tuukka (2018) Populist knowledge: 'Post-truth' repertoires of contesting epistemic authorities. European Journal of Cultural and Political Sociology, 5(4), 1–33.

Ylä-Anttila, Tuukka, Eranti, Veikko, Kukkonen, Anna (2018) Aihemallinnuksesta kehysmallinnukseen. Politiikka, 60(2), 148–156.

Yu, Bei, Kaufmann, Stefan, Diermeier, Daniel (2008) Classifying Party Affiliation from Political Speech. Journal of Information Technology. Politics, 5(1), 33–48.

Appendix

Appendix 1: The process of LDA

The process of the LDA-application is iterative in its three levels. First, the computer assigns each document with a topic distribution based on dirichlet distribution, i.e. it randomly (although controlled by a prior called α) chooses how much a document is about each topic. In the retirement discussion -corpus, there were 135 992 documents, from which it was decided to find 200 topics. Because the initial α in the thesis' application was symmetric, it allocated each document a probability of 1/200 to belong to each of the topics. During the iterations, the model learns asymmetric α from the data, i.e. it corrects its estimation in which proportions each topic is presented in each of the documents. The non-symmetricity of the α enabled the short conversations pondering over only few issues to have small probabilities to belong to most of the topics, and larger probabilities to only a couple of the topics; a necessary feature for this kind of corpus where the length of the units of analysis varies. (Rehurek & Sojka 2010.)

Before starting the iterations, the model had two more distributions to allocate. Next, it randomly (although again controlled by a dirichlet prior, η this time) assigned each word in each document a probability to belong into any of the 200 topics within the document. At this stage, the model had no idea about how the words in other documents were allocated. The third step was to take a holistic picture of all the topic and word allocations in all the documents in the corpus and create the corpus-wide 200 topics, i.e. 200 groups of words, where each word in the corpus is allocated a probability to belong to each of the corpus-wide topics based on the previous two steps.

Then the model started iterating, evaluating and correcting the distributions. It went through 2000 documents at a time, iterating the within-document distribution of the topics 1000 times or less if the change in the topical distribution converged to less than 0.001[8]. After each 2000 documents, it evaluated the corpus-wide topic-distributions and output an example of five topics as well as a numerical value of the differences between the previous and the current topics. For the evaluations, it used Gibbs sampling to analyse the correctness of the topics: if word w in document d had a high probability to belong to, let's say, topic number 3 within the whole corpus, and the

[8] I also tried 5000 iterations, but, although this was not included in the information the model printed, apparently the convergence threshold of 0.001 was always reached within 1000 iterations as the results of the two models were identical and took equal amount of time.

topic number 3 had a high prevalence in the document d, it was likely that the word w was assigned to belong to topic number 3 in the document d. After going through the first 2000 documents, the change from the initial topics to the updated topics was 195.967380, on a scale from infinity to 0. After the next 2000 documents, the difference had already dropped to 9.766355, from which it continued to slowly converge towards 0. Already when the model was halfway going through the documents, the difference was approximately 0.6, which was not much different to the 0.541286 when it had evaluated all the 135 992 documents. In Gensim, it is possible to set the model to go through the corpus multiple times, but every pass through increases the time the model runs. As the difference between the updates was converging already in a slow pace, and the changes were already below one, I considered one pass a good enough result. In many libraries, it is not even possible to run the model through multiple passes. (Blei, Ng, Jordan 2003, Levy& Franklin 2014, Doig 2015, Griffiths & Steyvers 2004).

Appendix 2: The codes

The code used in the thesis is available at https://github.com/tirri/thesis

Appendix 3: Model output

[(0, ’0.092*“vaimo“ + 0.078*“kaveri“ + 0.061*“ystävä“ + 0.037*“tutustua“ + 0.031*“harrastaa“ + 0.026*“hevonen“ + 0.024*“hauska“ + 0.023*“kertoa“ + 0.021*“tietää“ + 0.020*“muistella“ + 0.019*“jalkapallo“ + 0.017*“jutella“ + 0.017*“vanha“ + 0.016*“tykätä“ + 0.013*“mieli“’),
1, ’0.070*“vaivata“ + 0.070*“luultavasti“ + 0.060*“oleminen“ + 0.050*“ehkäpä“ + 0.049*“vaja“ + 0.048*“uskollinen“ + 0.045*“otin“ + 0.044*“kiltti“ + 0.042*“sakko“ + 0.023*“sanoa“ + 0.017*“rakastaa“ + 0.017*“sairaat#eläkeläinen“ + 0.012*“lausahdus“ + 0.011*“kuusi#mäki“ + 0.011*“haapa#salo“’),
2, ’0.188*“niinistö“ + 0.104*“sota“ + 0.092*“usa“ + 0.068*“nato“ + 0.046*“media“ + 0.040*“suomi“ + 0.023*“hyökkäys“ + 0.019*“sauli“ + 0.019*“monikulttuurisuus“ + 0.016*“syrjäytyä“ + 0.016*“nälkä“ + 0.016*“housut“ + 0.015*“valinta“ + 0.013*“fbi“ + 0.012*“toisin“’),
3, ’0.029*“pärjätä“ + 0.025*“hankkia“ + 0.024*“tehdä“ + 0.024*“raha“ + 0.017*“tarvita“ + 0.016*“elää“ + 0.013*“riittää“ + 0.013*“käydä“ + 0.012*“vaate“ + 0.012*“kulku“ + 0.012*“koto“ + 0.011*“pari“ + 0.011*“pitää“ + 0.011*“kunto“ + 0.011*“kämppä“’),

4, ’0.097*“asunto“ + 0.080*“hinta“ + 0.068*“ostaa“ + 0.033*“vuokra“ + 0.029*“maksaa“ + 0.026*“halpa“ + 0.024*“raha“ + 0.023*“kallis“ + 0.021*“myydä“ + 0.017*“asua“ + 0.012*“muuttaa“ + 0.012*“maksa“ + 0.012*“asu“ + 0.012*“elää“ + 0.011*“pitää“’),

5, ’0.099*“tasa“ + 0.055*“arvo“ + 0.048*“vastustaa“ + 0.041*“kieli“ + 0.029*“rikosilmoitus“ + 0.027*“suojella“ + 0.023*“suomi“ + 0.022*“vähemmistö“ + 0.018*“rkp“ + 0.018*“konklaavi“ + 0.014*“commissio“ + 0.014*“maittaa“ + 0.014*“ihmisarvo“ + 0.013*“rekisteröidä“ + 0.013*“kieliä“’),

6, ’0.044*“johtaja“ + 0.042*“laki“ + 0.039*“oikeus“ + 0.031*“virka“ + 0.026*“jäsen“ + 0.025*“tehtävä“ + 0.023*“virkamies“ + 0.023*“ylijohtaja“ + 0.022*“viranomainen“ + 0.020*“hallinto“ + 0.019*“päätös“ + 0.017*“korkea“ + 0.015*“kansliapäällikkö“ + 0.013*“johto“ + 0.013*“kansa#eläke#laitos“’),

7, ’0.113*“syödä“ + 0.103*“lääke“ + 0.026*“täydellisyys“ + 0.025*“määrätä“ + 0.022*“itsemurha“ + 0.018*“osua“ + 0.016*“käyttää“ + 0.016*“kevyt“ + 0.016*“tausta“ + 0.015*“mielestä“ + 0.014*“ajautua“ + 0.014*“kokoinen“ + 0.013*“lopettaa“ + 0.013*“syöttää“ + 0.012*“ohitse“’),

8, ’0.045*“jalka“ + 0.031*“työ#voima#pula“ + 0.019*“kävellä“ + 0.018*“käsi“ + 0.017*“alkaa“ + 0.017*“piste“ + 0.016*“pitää“ + 0.011*“laittaa“ + 0.011*“käydä“ + 0.011*“haitata“ + 0.010*“näyttää“ + 0.009*“ottaa“ + 0.009*“testata“ + 0.008*“näkö“ + 0.008*“sanoa“’),

9, ’0.115*“rakkaus“ + 0.047*“koti“ + 0.046*“uusia“ + 0.041*“rauha“ + 0.039*“täydellinen“ + 0.031*“lähimmäinen“ + 0.030*“yksilö“ + 0.029*“varoitus“ + 0.028*“irti“ + 0.025*“turva“ + 0.019*“petos“ + 0.018*“muutos#voima“ + 0.018*“maa“ + 0.017*“isänmaa“ + 0.017*“sielu“’),

10, ’0.051*“musta“ + 0.043*“kuolema“ + 0.036*“pikku“ + 0.027*“vankila“ + 0.024*“valehdella“ + 0.023*“lentää“ + 0.023*“metri“ + 0.021*“turvallinen“ + 0.020*“vanhuus“ + 0.020*“kehdata“ + 0.020*“pohja“ + 0.019*“linna“ + 0.018*“pinta“ + 0.018*“henkistää“ + 0.017*“vanki“’),

11, ’0.121*“maa“ + 0.117*“eu“ + 0.076*“suomia“ + 0.046*“maailma“ + 0.042*“suomi“ + 0.039*“euroopa“ + 0.030*“venäjä“ + 0.021*“valta“ + 0.019*“kansa“ + 0.014*“tuho“ + 0.012*“hallita“ + 0.012*“taistelu“ + 0.011*“kiina“ + 0.011*“-rakentaminen“ + 0.011*“ihmiskunta“’),

12, ’0.090*“lehti“ + 0.041*“kanslia“ + 0.039*“oy“ + 0.032*“oikeusasiamies“ + 0.028*“arvostella“ + 0.028*“tilaisuus“ + 0.022*“islam“ + 0.020*“lainaus“ + 0.019*“lääkäri#jäsen“ + 0.019*“ammattilainen“ + 0.017*“menettely“ + 0.017*“kehotus“ + 0.016*“vm“ + 0.015*“kirjallinen“ + 0.015*“työkokemus“’),

13, ’0.128*“henkilö“ + 0.030*“tieto“ + 0.028*“kyseinen“ + 0.022*“toimenpide“ + 0.021*“tarkka“ + 0.020*“laki“ + 0.019*“tarkoittaa“ + 0.018*“mukainen“ + 0.016*“syy“ + 0.016*“työvoimatoimisto“ + 0.016*“tällainen“ + 0.016*“edellyttää“ + 0.015*“taho“ + 0.015*“oikeus“ + 0.014*“kirjata“’),

14, ’0.056*“yliopisto“ + 0.037*“valmistua“ + 0.034*“huolenpito“ + 0.030*“saudi“ + 0.029*“opinnot“ + 0.022*“morsian“ + 0.018*“opiskella“ + 0.018*“terroristi“ + 0.014*“armoittaa“ + 0.014*“sitra“ + 0.013*“kavaltaa“ + 0.012*“presidenttiehdokas“ + 0.011*“päämies“ + 0.010*“loukkaava“ + 0.009*“maisteri“’),

15, ’0.304*“auto“ + 0.044*“vaihtaa“ + 0.030*“ostaa“ + 0.028*“omistaja“ + 0.025*“vaihto“ + 0.025*“kaupallinen“ + 0.020*“vika“ + 0.018*“käyttää“ + 0.016*“tammi“ + 0.015*“km“ + 0.012*“moottori“ + 0.011*“alapuoli“ + 0.010*“ostaja“ + 0.010*“remontti“ + 0.009*“tkm“’),

16, ’0.250*“kirjoittaa“ + 0.095*“teko“ + 0.052*“päätös“ + 0.040*“vakuutusoikeus“ + 0.029*“minkälainen“ + 0.024*“joutaa“ + 0.023*“allekirjoittaa“ + 0.017*“reilusti“ + 0.014*“yksimielisyys“ + 0.014*“kiinnostaa“ + 0.012*“kuja“ + 0.012*“lausuma“ + 0.012*“keskeyttää“ + 0.011*“huikea“ + 0.011*“voimallinen“’),

17, ’0.066*“aine“ + 0.053*“eliitti“ + 0.041*“avustus“ + 0.031*“katto“ + 0.031*“luopua“ + 0.027*“raaka“ + 0.026*“paluu“ + 0.026*“järvi“ + 0.024*“vievä“ + 0.021*“omistus“ + 0.020*“ruua“ + 0.020*“sankari“ + 0.018*“maa“ + 0.018*“metsänomistaja“ + 0.017*“tasainen“’),

18, ’0.107*“mukaisesti“ + 0.064*“itä“ + 0.051*“kasvot“ + 0.047*“vaikeus“ + 0.046*“israeli“ + 0.032*“puolustus“ + 0.029*“kuljettaa“ + 0.028*“säätiö“ + 0.026*“koota“ + 0.024*“tietoinen“ + 0.024*“luopuminen“ + 0.023*“valuutta“ + 0.022*“synty“ + 0.020*“lukumäärä“ + 0.018*“nordea“’),

19, ’0.124*“muutos“ + 0.066*“puute“ + 0.051*“yhteisö“ + 0.049*“taito“ + 0.044*“vailla“ + 0.041*“muutto“ + 0.033*“säilyttää“ + 0.032*“periaate“ + 0.027*“arvostus“ + 0.025*“aikataulu“ + 0.024*“vaari“ + 0.022*“hetkinen“ + 0.021*“nikula“ + 0.021*“kaltainen“ + 0.017*“pätevyys“’),

20, ’0.110*“köyhä“ + 0.058*“kokoomus“ + 0.053*“rikas“ + 0.030*“raha“ + 0.028*“eläkeläinen“ + 0.023*“suomi“ + 0.019*“työtön“ + 0.016*“kansa“ + 0.016*“pitää“ + 0.014*“elää“ + 0.012*“ihminen“ + 0.012*“tavallinen“ + 0.011*“köyhyys“ + 0.009*“opiskelija“ + 0.009*“lapsiperhe“’),

21, ’0.146*“ruotsi“ + 0.079*“kunnia“ + 0.044*“säästyä“ + 0.040*“palvelus“ + 0.038*“muttei“ + 0.038*“lypsää“ + 0.038*“tippua“ + 0.036*“seksuaalinen“ + 0.031*“orja“ + 0.023*“tuottaja“ + 0.023*“konsti“ + 0.019*“suomi“ + 0.018*“kurittaa“ + 0.017*“hyväksikäyttö“ + 0.015*“karsia“’),

22, ’0.142*“ohjelma“ + 0.115*“tv“ + 0.068*“kulttuuri“ + 0.038*“radio“ + 0.038*“lähestyä“ + 0.036*“ssä“ + 0.026*“ylen“ + 0.019*“suo“ + 0.019*“innostua“ + 0.018*“haluttaa“ + 0.018*“uskonsuunta“ + 0.018*“itsekkyys“ + 0.017*“epämääräinen“ + 0.015*“versio“ + 0.014*“kertoma“’),

23, ’0.306*“eläkeläinen“ + 0.085*“sdp“ + 0.085*“demari“ + 0.035*“luvata“ + 0.017*“lupaus“ + 0.017*“pitää“ + 0.015*“kokoomus“ + 0.013*“tehdä“ + 0.012*“vaalia“ + 0.010*“muistaa“ + 0.010*“taitaa“ + 0.009*“nostaa“ + 0.009*“näyttää“ + 0.008*“ben“ + 0.008*“unohtaa“’),

24, ’0.100*“pistää“ + 0.065*“itsenäinen“ + 0.052*“ihana“ + 0.044*“suositella“ + 0.035*“kovasti“ + 0.029*“tarvia“ + 0.028*“kauhea“ + 0.027*“pelle“ + 0.022*“tarvi“ + 0.020*“pätkä“ + 0.020*“sopeutua“ + 0.018*“moittia“ + 0.017*“vitsaus“ + 0.017*“tasavertainen“ + 0.017*“naurattaa“’),

25, ’0.059*“matka“ + 0.049*“paikka“ + 0.044*“viikko“ + 0.042*“tunti“ + 0.042*“käydä“ + 0.030*“mukava“ + 0.026*“pari“ + 0.021*“päivä“ + 0.019*“hotelli“ + 0.019*“eläkeläinen“ + 0.017*“löytyä“ + 0.014*“lippu“ + 0.014*“päästä“ + 0.013*“pitkä“ + 0.011*“kesä“’),

26, ’0.327*“valtio“ + 0.186*“kunta“ + 0.049*“miljoona“ + 0.044*“sonera“ + 0.023*“miljardi“ + 0.016*“siirtää“ + 0.013*“ottaa“ + 0.012*“henkilökunta“ + 0.010*“palkkaus“ + 0.010*“kuolintodistus“ + 0.010*“suomi“ + 0.009*“maksettava“ + 0.009*“virkamies“ + 0.008*“päärakennus“ + 0.008*“syrjäyttää“’),

27, ’0.127*“noudattaa“ + 0.063*“kyetä“ + 0.062*“lopullinen“ + 0.055*“amerikka“ + 0.043*“niittää“ + 0.033*“tyhjentää“ + 0.032*“tuuli“ + 0.025*“rahasto“ + 0.025*“irrota“ + 0.024*“kieltäytyä“ + 0.023*“kyseenalaistaa“ + 0.019*“hiljaisuus“ + 0.019*“virkailija“ + 0.018*“osoittautua“ + 0.017*“kynsi“’),

28, ’0.170*“koira“ + 0.063*“juoda“ + 0.033*“naama“ + 0.031*“ominaisuus“ + 0.025*“kertomus“ + 0.020*“haaveilla“ + 0.020*“hypätä“ + 0.020*“tuuma“ + 0.019*“viina“ + 0.017*“mökki“ + 0.017*“väsyä“ + 0.016*“ottaa“ + 0.014*“härski“ + 0.014*“maku“ + 0.014*“mökä“’),

29, ’0.054*“helvetti“ + 0.040*“kylä“ + 0.036*“tyyli“ + 0.033*“menetelmä“ + 0.029*“kaikenlainen“ + 0.027*“esimies“ + 0.027*“vene“ + 0.025*“seisoa“ + 0.024*“hiekka“ + 0.022*“viimein“ + 0.019*“luki“ + 0.018*“ympärille“ + 0.015*“vammautua“ + 0.015*“tori“ + 0.014*“suunnaton“’),

30, ’0.064*“käsitys“ + 0.051*“lainata“ + 0.045*“millainen“ + 0.044*“toistensa“ + 0.038*“välittömästi“ + 0.037*“tyypillinen“ + 0.036*“tarina“ + 0.028*“hun“ + 0.026*“erikoinen“ + 0.022*“katsoma“ + 0.021*“helmikuu“ + 0.021*“kuuma“ + 0.021*“unelma“ + 0.019*“kas“ + 0.018*“dokumentti“’),

31, ’0.060*“perintö“ + 0.034*“urheilu“ + 0.029*“kohtuuton“ + 0.028*“jakaja“ + 0.028*“kuluttaja“ + 0.026*“huvittaa“ + 0.023*“lukuunottaa“ + 0.021*“mafia“ + 0.020*“kohdistaa“ + 0.019*“kolkutella“ + 0.019*“poimia“ + 0.016*“perillinen“ + 0.015*“hyväkuntoinen“ + 0.015*“useiden“ + 0.014*“valvoa“’),
32, ’0.096*“tuollainen“ + 0.047*“toista“ + 0.039*“yleisö“ + 0.036*“laulaa“ + 0.033*“bensa“ + 0.030*“huomattavasti“ + 0.026*“törmätä“ + 0.026*“vastakkainasettelu“ + 0.025*“positiivinen“ + 0.021*“näyttö“ + 0.021*“kaatopaikka“ + 0.015*“penni“ + 0.015*“tapainen“ + 0.014*“ajokortti“ + 0.014*“suojatyöpaikka“’),
33, ’0.150*“hyvinvointi“ + 0.103*“talous“ + 0.057*“työvoima“ + 0.032*“hyvinvointivaltio“ + 0.031*“työllisyys“ + 0.030*“haaste“ + 0.029*“kääntyä“ + 0.025*“-rakentaminen“ + 0.023*“kokonaisuus“ + 0.022*“kotimainen“ + 0.016*“samanaikaisesti“ + 0.015*“malli“ + 0.015*“vastine“ + 0.013*“nykypäivä“ + 0.012*“työikäinen“’),
34, ’0.240*“herra“ + 0.042*“liikenne“ + 0.023*“tasava“ + 0.022*“pysähtyä“ + 0.022*“taakse“ + 0.022*“ruveta“ + 0.021*“apostolinen“ + 0.020*“lakko“ + 0.019*“eläke#raha“ + 0.017*“lukeutua“ + 0.014*“tahti“ + 0.013*“kaaos“ + 0.012*“vaientaa“ + 0.012*“enne#kui“ + 0.011*“alkaa“’),
35, ’0.142*“työpaikka“ + 0.087*“työ“ + 0.056*“tehdä“ + 0.052*“pitää“ + 0.035*“sossu“ + 0.027*“ihminen“ + 0.025*“systeemi“ + 0.024*“perse“ + 0.020*“välinpitämättömyys“ + 0.019*“työllistää“ + 0.015*“pois“ + 0.015*“raha“ + 0.014*“riittää“ + 0.013*“päästä“ + 0.013*“ottaa“’),
36, ’0.114*“numero“ + 0.040*“perehtyä“ + 0.035*“moraalittomuus“ + 0.034*“suosio“ + 0.033*“halukas“ + 0.030*“teuvo“ + 0.027*“mika“ + 0.025*“ohjus“ + 0.024*“kake“ + 0.023*“vaaliehdokas“ + 0.021*“lakialoite“ + 0.020*“autio“ + 0.020*“häkkinen“ + 0.019*“ohjaaminen“ + 0.018*“väittely“’),
37, ’0.101*“oppi“ + 0.074*“asema“ + 0.068*“ohjata“ + 0.064*“astua“ + 0.059*“ulkomaalainen“ + 0.042*“pakolainen“ + 0.036*“muuttaminen“ + 0.028*“valitsija“ + 0.028*“selkeä“ + 0.024*“liittää“ + 0.023*“rasisti“ + 0.021*“apulainen“ + 0.019*“erillinen“ + 0.018*“alkuperäinen“ + 0.017*“sotaveteraani“’),
38, ’0.049*“blondi“ + 0.034*“rauhallinen“ + 0.032*“selkä“ + 0.029*“kipu“ + 0.023*“mieliä“ + 0.023*“saavutus“ + 0.022*“työttömyyseläke“ + 0.021*“internet“ + 0.018*“härkä“ + 0.018*“päästä“ + 0.017*“hai“ + 0.013*“mahdoton“ + 0.012*“turvautua“ + 0.012*“hermo“ + 0.011*“etsiä“’),
39, ’0.066*“ihminen“ + 0.028*“yhteiskunta“ + 0.020*“seuraus“ + 0.019*“armo“ + 0.017*“todellisuus“ + 0.014*“ongelma“ + 0.014*“pitää“ + 0.013*“tulevaisuus“ +

0.013*“tarjolla“ + 0.013*“tehdä“ + 0.011*“elää“ + 0.011*“maailma“ + 0.011*“alkaa“ + 0.010*“ratkaista“ + 0.010*“johtaa“’),
40, ’0.072*“aktiivinen“ + 0.057*“kehittäminen“ + 0.045*“varmistaa“ + 0.044*“kansalaispalkka“ + 0.041*“nähden“ + 0.031*“neuvoja“ + 0.027*“teknologia“ + 0.026*“ajaminen“ + 0.025*“katketa“ + 0.023*“ulkopolitiikka“ + 0.023*“rikkoa“ + 0.023*“toimiala“ + 0.022*“toimija“ + 0.022*“hölmö“ + 0.021*“merkitys“’),
41, ’0.116*“äänestää“ + 0.077*“vihreä“ + 0.055*“duunari“ + 0.042*“puolustaja“ + 0.037*“vasemmisto“ + 0.031*“puolustaa“ + 0.027*“oikeastaan“ + 0.026*“mukautua“ + 0.025*“kurjuus“ + 0.020*“etu“ + 0.019*“ääni“ + 0.018*“voittaja“ + 0.016*“pitää“ + 0.014*“tehdä“ + 0.012*“valta“’),
42, ’0.058*“päteä“ + 0.040*“asukas“ + 0.040*“järkevä“ + 0.039*“pula“ + 0.032*“paikata“ + 0.026*“vapautua“ + 0.026*“taksi“ + 0.026*“halva“ + 0.025*“ohella“ + 0.020*“uida“ + 0.019*“taikka“ + 0.019*“paikka“ + 0.019*“avustaja“ + 0.018*“vähänkin“ + 0.014*“kysyntä“’),
43, ’0.087*“suu“ + 0.067*“ruoka“ + 0.067*“vesi“ + 0.052*“istua“ + 0.043*“täynnä“ + 0.031*“paska“ + 0.026*“käydä“ + 0.025*“katua“ + 0.020*“äkkiä“ + 0.017*“tervehtiä“ + 0.016*“työntää“ + 0.015*“ostos“ + 0.015*“lasi“ + 0.014*“tuhoava“ + 0.011*“tie“’),
44, ’0.080*“pahoinvointi“ + 0.058*“kissa“ + 0.056*“eläin“ + 0.048*“seikka“ + 0.034*“hävittää“ + 0.033*“turkki“ + 0.032*“suoja“ + 0.030*“nauru“ + 0.028*“kortisto“ + 0.027*“rotu“ + 0.027*“lähivuosi“ + 0.023*“viitata“ + 0.022*“ruokkia“ + 0.021*“siipi“ + 0.020*“sais“’),
45, ’0.054*“vapaus“ + 0.050*“tyytyä“ + 0.046*“ilo“ + 0.038*“koto“ + 0.033*“askel“ + 0.027*“onnellinen“ + 0.025*“ikuinen“ + 0.024*“aurinko“ + 0.021*“elo“ + 0.019*“ympärillä“ + 0.016*“pala“ + 0.015*“meikäläinen“ + 0.014*“koittaa“ + 0.014*“kulua“ + 0.014*“näkymä“’),
46, ’0.063*“tila“ + 0.057*“totuus“ + 0.055*“vastuu“ + 0.052*“demokratia“ + 0.046*“mahdollisuus“ + 0.044*“rakentaa“ + 0.042*“valta“ + 0.032*“terve“ + 0.029*“luottaa“ + 0.027*“oikeus“ + 0.027*“virhe“ + 0.025*“väärä“ + 0.025*“tahto“ + 0.025*“korjata“ + 0.022*“ihminen“’),
47, ’0.091*“katu“ + 0.076*“pimeä“ + 0.047*“suku“ + 0.035*“semmonen“ + 0.033*“vartija“ + 0.026*“minul“ + 0.025*“tyttöystävä“ + 0.025*“ensiksi“ + 0.025*“nurkka“ + 0.024*“mädännäisyys“ + 0.023*“pittää“ + 0.023*“sokeasti“ + 0.022*“potkia“ + 0.021*“olut“ + 0.019*“ny“’),
48, ’0.168*“helsinki“ + 0.052*“ulko-#asia#neuvos“ + 0.037*“tampere“ + 0.031*“home“ + 0.028*“hautajaiset“ + 0.023*“kallio“ + 0.022*“oulu“ + 0.021*“keskustalainen“ + 0.021*“jyvänen#kylä“ + 0.019*“espoo“ +

0.017*“tutkimuskeskus“ + 0.017*“valmistautua“ + 0.016*“joulukuu“ +
0.016*“todennäköinen“ + 0.016*“suostumus“’),
49, ’0.058*“ps“ + 0.052*“ajella“ + 0.049*“mersu“ + 0.037*“hommata“ +
0.037*“virta“ + 0.035*“mainostaa“ + 0.031*“ovi“ + 0.028*“kirous“ + 0.024*“kehua“ +
0.024*“temppu“ + 0.023*“polttoaine“ + 0.019*“kollega“ + 0.018*“ritva“ +
0.017*“poistua“ + 0.016*“kauris“’),
50, ’0.079*“pääjohtaja“ + 0.077*“vanhanen“ + 0.072*“matti“ + 0.049*“hallitus“ +
0.047*“pääministeri“ + 0.037*“kekkonen“ + 0.035*“katainen“ + 0.024*“suomi“ +
0.023*“puheenjohtaja“ + 0.019*“jaakko“ + 0.017*“jyrki“ + 0.017*“vanhan“ +
0.017*“heinä#luoma“ + 0.015*“jorma“ + 0.014*“ministeri“’),
51, ’0.034*“tuomita“ + 0.032*“tuomio“ + 0.029*“lakimies“ + 0.025*“tuomari“ +
0.024*“teko“ + 0.024*“todiste“ + 0.023*“petos“ + 0.023*“käyttää“ + 0.023*“todeta“ +
0.020*“väite“ + 0.018*“vakava“ + 0.018*“aiheuttaa“ + 0.016*“puolueeton“ +
0.016*“oikeus“ + 0.015*“viha“’),
52, ’0.128*“lasku“ + 0.071*“tili“ + 0.057*“rangaistus“ + 0.057*“raha“ +
0.032*“maahanmuuttaja“ + 0.029*“sähköposti“ + 0.028*“viimeistään“ +
0.024*“työtoveri“ + 0.024*“suunnata“ + 0.018*“bkt“ + 0.016*“harjoitella“ +
0.014*“muutoinkin“ + 0.012*“koetella“ + 0.012*“työttömyys#prosentti“ +
0.011*“käteinen“’),
53, ’0.122*“presidentti“ + 0.100*“halonen“ + 0.046*“kirja“ + 0.035*“tarja“ +
0.024*“vanhurskas“ + 0.022*“saulia“ + 0.021*“koivisto“ + 0.019*“sanoma“ +
0.015*“ahti#saari“ + 0.014*“bush“ + 0.010*“valita“ + 0.007*“toveri“ + 0.007*“qukoa“ +
0.007*“ulkoministeri“ + 0.007*“jumalaton“’),
54, ’0.033*“kasvaa“ + 0.032*“neuvottelu“ + 0.027*“irtisanoa“ + 0.024*“työntekijä“ +
0.023*“yt“ + 0.021*“tuotanto“ + 0.019*“lama“ + 0.016*“henkilöstö“ +
0.013*“vähentää“ + 0.013*“suomi“ + 0.012*“oyj“ + 0.012*“kysyntä“ +
0.011*“kansantalous“ + 0.011*“työpaikka“ + 0.011*“veronalennus“’),
55, ’0.179*“äiti“ + 0.165*“isä“ + 0.075*“veli“ + 0.049*“rakennus“ + 0.029*“varoittaa“
+ 0.020*“kuolla“ + 0.018*“pitkin“ + 0.017*“syntymä“ + 0.017*“vanha“ +
0.016*“musiikki“ + 0.015*“vaiva“ + 0.014*“tuoja“ + 0.013*“surullinen“ +
0.012*“totisesti“ + 0.011*“pettyä“’),
56, ’0.204*“esko“ + 0.033*“suunnitelma“ + 0.026*“pohtia“ + 0.020*“neuvotella“ +
0.017*“pohja“ + 0.017*“suunnitella“ + 0.015*“haitta“ + 0.012*“välinen“ +
0.012*“väistää“ + 0.012*“enin“ + 0.011*“laajentaa“ + 0.011*“alueellinen“ +
0.011*“työllistäminen“ + 0.010*“vaihtoehto“ + 0.010*“liittyä“’),
57, ’0.149*“ääni“ + 0.142*“odottaa“ + 0.063*“päästää“ + 0.050*“kiire“ +
0.042*“pääsy“ + 0.040*“päästä“ + 0.038*“valita“ + 0.023*“vihta“ +

0.021*“myönteinen“ + 0.018*“eläke#päätös“ + 0.017*“pelko“ + 0.017*“aktiivisesti“ + 0.016*“jatko“ + 0.015*“vaivainen“ + 0.012*“tuore“’),
58, ’0.141*“kuukausi“ + 0.111*“täysi“ + 0.067*“tienata“ + 0.063*“mahtaa“ + 0.038*“työaika“ + 0.036*“päivä“ + 0.036*“palkka“ + 0.031*“tuhlata“ + 0.027*“viikko“ + 0.019*“elää“ + 0.019*“kk“ + 0.015*“työ“ + 0.014*“varoitella“ + 0.014*“pv“ + 0.013*“elikkä“’),
59, ’0.083*“muistuttaa“ + 0.041*“verottaja“ + 0.041*“tiukka“ + 0.039*“palaute“ + 0.035*“rikollisuus“ + 0.031*“jollei“ + 0.030*“erinomainen“ + 0.030*“aikaisemmin“ + 0.024*“ara#järvi“ + 0.023*“bisnes“ + 0.021*“luettaa“ + 0.021*“toivottaa“ + 0.020*“eilinen“ + 0.017*“taival“ + 0.016*“materiaali“’),
60, ’0.056*“muille“ + 0.052*“kaikkinainen“ + 0.051*“virasto“ + 0.034*“eläkeikäinen“ + 0.032*“oikeuslaitos“ + 0.030*“juoppo“ + 0.030*“arvata“ + 0.026*“arvoinen“ + 0.023*“motto“ + 0.023*“orjuus“ + 0.022*“loinen“ + 0.020*“parhaaksi“ + 0.020*“eläke#kertymä“ + 0.019*“käytäntö“ + 0.018*“egyptin“’),
61, ’0.177*“kauppa“ + 0.066*“kumma“ + 0.060*“luokka“ + 0.046*“hyssälä“ + 0.037*“teollisuusministeriö“ + 0.029*“soida“ + 0.025*“neuvonantaja“ + 0.021*“pukeutua“ + 0.020*“pöytäkirja“ + 0.018*“tunnetusti“ + 0.015*“loputon“ + 0.014*“nettotulo“ + 0.012*“parhaimmat“ + 0.012*“kesäkuu“ + 0.011*“mutka“’),
62, ’0.300*“maksaa“ + 0.067*“korvaus“ + 0.046*“veronmaksaja“ + 0.034*“maksa“ + 0.022*“raha“ + 0.020*“maksaminen“ + 0.020*“korvata“ + 0.020*“valtio“ + 0.018*“tapaturma“ + 0.016*“kustannus“ + 0.013*“palkka“ + 0.013*“verovarat“ + 0.012*“kallis“ + 0.012*“suuruinen“ + 0.011*“eläkemaksu“’),
63, ’0.075*“hoito“ + 0.061*“sairaus“ + 0.043*“masennus“ + 0.032*“ongelma“ + 0.031*“oire“ + 0.020*“läheinen“ + 0.019*“masentua“ + 0.018*“työryhmä“ + 0.018*“auttaa“ + 0.015*“katkera“ + 0.014*“syy“ + 0.013*“heimo“ + 0.012*“hoitaa“ + 0.012*“johtua“ + 0.012*“tietää“’),
64, ’0.076*“raja“ + 0.051*“maaseutu“ + 0.047*“kaivata“ + 0.037*“suola“ + 0.029*“tarpeellinen“ + 0.029*“sali“ + 0.027*“pilvi“ + 0.025*“melkoinen“ + 0.020*“hyvin“ + 0.020*“homo“ + 0.019*“pesä“ + 0.019*“kunnioitus“ + 0.019*“kuula“ + 0.017*“suojella“ + 0.016*“hukkua“’),
65, ’0.239*“klo“ + 0.124*“kysymys“ + 0.080*“vastaus“ + 0.050*“heh“ + 0.039*“vähentää“ + 0.037*“puh“ + 0.030*“terveiset“ + 0.028*“vähennys“ + 0.018*“tarkemmin“ + 0.017*“tulo“ + 0.015*“aiheutua“ + 0.014*“ansiotulo“ + 0.013*“ottaa“ + 0.012*“saastuttaa“ + 0.011*“kysyä“’),
66, ’0.107*“kuningas“ + 0.077*“velka“ + 0.048*“idea“ + 0.039*“alivaltiosihteeri“ + 0.039*“lisääntyä“ + 0.029*“valmistaa“ + 0.027*“liikaa“ + 0.026*“vähetä“ +

0.025*“kehittyä“ + 0.024*“viina“ + 0.023*“muilla“ + 0.023*“sietää“ + 0.019*“täky“ + 0.017*“koskettaa“ + 0.017*“kasvaa“’),
67, ’0.098*“kokous“ + 0.066*“ilmainen“ + 0.051*“saaja“ + 0.040*“kirje“ + 0.035*“peru“ + 0.033*“osto“ + 0.033*“ylimääräinen“ + 0.033*“vaate“ + 0.027*“maaliskuu“ + 0.025*“eräänlainen“ + 0.022*“läsnä“ + 0.021*“osakas“ + 0.018*“oppilaitos“ + 0.017*“kultainen“ + 0.015*“valtuusto“’),
68, ’0.144*“palvelu“ + 0.106*“yhteinen“ + 0.038*“turvallisuus“ + 0.027*“uudistus“ + 0.020*“teema“ + 0.019*“osapuoli“ + 0.017*“luoma“ + 0.017*“päivähoito“ + 0.016*“sosiaali“ + 0.016*“resurssit“ + 0.015*“tärkeä“ + 0.015*“osittain“ + 0.015*“rajata“ + 0.014*“sokeri“ + 0.014*“ratkaisu“’),
69, ’0.171*“koulutus“ + 0.077*“ammatti“ + 0.068*“opiskella“ + 0.039*“työelämä“ + 0.033*“tutkinto“ + 0.025*“vakituinen“ + 0.025*“harrastus“ + 0.024*“kilpailu“ + 0.022*“alkoholisti“ + 0.017*“adventtikirkko“ + 0.013*“suuttua“ + 0.011*“käydä“ + 0.011*“suorittaminen“ + 0.011*“palkinto“ + 0.010*“adventisti“’),
70, ’0.098*“tähti“ + 0.077*“nauraa“ + 0.057*“hyöty“ + 0.048*“iskeä“ + 0.040*“polttaa“ + 0.040*“tupakka“ + 0.032*“paljoa“ + 0.028*“vaa“ + 0.027*“kotka“ + 0.026*“nimetä“ + 0.024*“vetää“ + 0.023*“työmies“ + 0.022*“vierailla“ + 0.021*“sänky“ + 0.020*“manninen“’),
71, ’0.066*“ihmeellinen“ + 0.056*“lapsiperhe“ + 0.052*“tietokone“ + 0.045*“yllätys“ + 0.042*“rivi“ + 0.041*“valtaosa“ + 0.034*“poikkeus“ + 0.029*“loistaa“ + 0.023*“viiva“ + 0.021*“jertyä“ + 0.020*“yksinkertainen“ + 0.020*“harkinta“ + 0.020*“todennäköisesti“ + 0.019*“pallo“ + 0.019*“tinkiä“’),
72, ’0.285*“hallitus“ + 0.073*“leikata“ + 0.056*“leikkaus“ + 0.031*“tehdä“ + 0.029*“lapsilisä“ + 0.029*“vaali“ + 0.018*“oppositio“ + 0.017*“päätös“ + 0.016*“politiikka“ + 0.015*“oikeisto“ + 0.014*“vasemmisto#liitto“ + 0.013*“pääministeri“ + 0.010*“rintama“ + 0.010*“hallituspuolue“ + 0.010*“jatkaa“’),
73, ’0.049*“ehdotus“ + 0.047*“kerto“ + 0.046*“riski“ + 0.037*“käyttäjä“ + 0.032*“verkosto“ + 0.030*“itsekäs“ + 0.028*“vasten“ + 0.026*“ilmi“ + 0.025*“yleisesti“ + 0.025*“kuulostaa“ + 0.021*“luokitella“ + 0.021*“häiriö“ + 0.020*“vetäytyä“ + 0.019*“ongelma“ + 0.019*“vaativa“’),
74, ’0.252*“nuori“ + 0.100*“vanha“ + 0.053*“porukka“ + 0.027*“eläkeläinen“ + 0.025*“pitää“ + 0.019*“päästä“ + 0.018*“tehdä“ + 0.016*“ryppy“ + 0.015*“jäädä“ + 0.015*“juosta“ + 0.014*“työ“ + 0.014*“pelastua“ + 0.013*“parasta“ + 0.012*“alkaa“ + 0.011*“taitaa“’),
75, ’0.054*“amp“ + 0.027*“juha“ + 0.026*“erkki“ + 0.025*“seppo“ + 0.021*“liisa“ + 0.020*“markku“ + 0.020*“antti“ + 0.020*“matti“ + 0.019*“heikki“ + 0.019*“timo“ + 0.019*“neiti“ + 0.016*“-mikko“ + 0.016*“arto“ + 0.016*“kari“ + 0.016*“raimo“’),

76, ’0.055*“komissio“ + 0.042*“oksa“ + 0.034*“sahata“ + 0.031*“johtava“ + 0.029*“viewtopic“ + 0.023*“vuosikausi“ + 0.022*“parlamentti“ + 0.021*“velallinen“ + 0.017*“luja“ + 0.016*“keskuspankki“ + 0.016*“out“ + 0.014*“ulottaa“ + 0.013*“riippumattomuus“ + 0.012*“jospa“ + 0.012*“vastustus“’),
77, ’0.063*“haukkua“ + 0.055*“hää“ + 0.050*“valtiovalta“ + 0.038*“vakavasti“ + 0.037*“uhka“ + 0.036*“luu“ + 0.030*“syyskuu“ + 0.030*“sairastua“ + 0.027*“siel“ + 0.024*“hylätä“ + 0.021*“säälittää“ + 0.021*“säännöllisesti“ + 0.017*“monenlainen“ + 0.016*“aika#eläke“ + 0.016*“kauaa“’),
78, ’0.054*“voitto“ + 0.050*“tuotto“ + 0.050*“yhtiö“ + 0.049*“osake“ + 0.048*“sijoittaa“ + 0.040*“omistaja“ + 0.038*“raha“ + 0.036*“optio“ + 0.031*“sijoitus“ + 0.029*“pääoma“ + 0.024*“amerikkalainen“ + 0.022*“yritys“ + 0.019*“sijoittaja“ + 0.019*“osinko“ + 0.018*“rahoittaa“’),
79, ’0.352*“kansa“ + 0.110*“lipponen“ + 0.057*“suomi“ + 0.040*“paavo“ + 0.031*“politiikka“ + 0.023*“vallata“ + 0.018*“perkele“ + 0.013*“mönkäre“ + 0.011*“sotilaallinen“ + 0.011*“paasi#kivi“ + 0.010*“ryöstää“ + 0.009*“paha“ + 0.009*“raiskata“ + 0.007*“sulka“ + 0.007*“saulin“’),
80, ’0.326*“työ“ + 0.155*“työtön“ + 0.107*“tehdä“ + 0.060*“palkka“ + 0.022*“palkata“ + 0.017*“homma“ + 0.016*“työntekijä“ + 0.015*“työpaikka“ + 0.014*“tarvita“ + 0.014*“elää“ + 0.011*“työttömyyskorvaus“ + 0.009*“päivä“ + 0.009*“raha“ + 0.007*“palkkatyö“ + 0.007*“päästä“’),
81, ’0.051*“laiminlyödä“ + 0.043*“esiintyä“ + 0.040*“al“ + 0.038*“sukupuoli“ + 0.031*“israel“ + 0.030*“arabia“ + 0.029*“saksalainen“ + 0.026*“lappi“ + 0.024*“haudata“ + 0.022*“luomakunta“ + 0.015*“käyttäminen“ + 0.015*“oikeusjärjestelmä“ + 0.015*“murhata“ + 0.014*“rahastaa“ + 0.014*“sitäpaitsi“’),
82, ’0.045*“sisään“ + 0.034*“tarkistaa“ + 0.033*“asiantunteva“ + 0.028*“pikkuhiljaa“ + 0.026*“komitea“ + 0.025*“niellä“ + 0.025*“juna“ + 0.023*“kiinnittää“ + 0.023*“lähetys“ + 0.022*“semmoinen“ + 0.021*“hyljeksiä“ + 0.019*“neuvosto“ + 0.017*“viemäri“ + 0.017*“velvollinen“ + 0.017*“huomio“’),
83, ’0.082*“verotus“ + 0.065*“vero“ + 0.034*“poistaa“ + 0.032*“oikeudenmukainen“ + 0.032*“tulo“ + 0.026*“pitää“ + 0.023*“oikeudenmukaisuus“ + 0.023*“verottaa“ + 0.019*“alentaa“ + 0.016*“nostaa“ + 0.014*“maksaa“ + 0.014*“malli“ + 0.014*“palkka“ + 0.014*“suurituloinen“ + 0.012*“keventää“’),
84, ’0.281*“suomi“ + 0.099*“suomalainen“ + 0.078*“maa“ + 0.034*“suomia“ + 0.030*“ulkomaa“ + 0.023*“saksa“ + 0.018*“muuttaa“ + 0.013*“espanja“ + 0.012*“kansalainen“ + 0.010*“ruotsi“ + 0.010*“euroopa“ + 0.009*“ranska“ + 0.009*“asua“ + 0.008*“eu“ + 0.007*“italia“’),

85, ’0.203*“nousta“ + 0.075*“nousu“ + 0.071*“korotus“ + 0.040*“taso“ + 0.037*“vuode“ + 0.035*“pelto“ + 0.032*“vuosikymmen“ + 0.022*“laiska“ + 0.019*“enemmän“ + 0.018*“duuni“ + 0.016*“tasapaino“ + 0.015*“kulukorvaus“ + 0.014*“konkurssi“ + 0.014*“kiihtyä“ + 0.013*“emäntä“’),

86, ’0.062*“jukka“ + 0.050*“ottaminen“ + 0.045*“sampo“ + 0.044*“sailas“ + 0.043*“huippu“ + 0.042*“toimitusjohtaja“ + 0.039*“harri“ + 0.039*“varma“ + 0.035*“risto“ + 0.033*“lilius“ + 0.030*“palkkio“ + 0.023*“todistaja“ + 0.021*“ilmarinen“ + 0.016*“mikael“ + 0.015*“optio“’),

87, ’0.077*“kuulla“ + 0.063*“jne“ + 0.048*“uskoa“ + 0.042*“pimeys“ + 0.041*“hullu“ + 0.040*“kulkea“ + 0.027*“vanhurskaus“ + 0.024*“uskova“ + 0.024*“pelätä“ + 0.023*“sanoa“ + 0.023*“kuski“ + 0.023*“nähdä“ + 0.020*“erä“ + 0.018*“bussi“ + 0.014*“pitäs“’),

88, ’0.103*“sairaala“ + 0.076*“romahtaa“ + 0.063*“kaatua“ + 0.052*“sisar“ + 0.049*“veteraani“ + 0.026*“isä“ + 0.024*“aisa“ + 0.019*“lyödä“ + 0.018*“naurettava“ + 0.016*“taloyhtiö“ + 0.016*“konserni“ + 0.015*“markka“ + 0.014*“pako“ + 0.013*“holkeri“ + 0.013*“roimasti“’),

89, ’0.075*“yhtiö“ + 0.046*“uutinen“ + 0.043*“kommentoida“ + 0.038*“julkaista“ + 0.033*“kertoa“ + 0.029*“hauta“ + 0.027*“laillinen“ + 0.025*“ämmä“ + 0.021*“suru“ + 0.019*“huora“ + 0.019*“järjestely“ + 0.019*“vieläpä“ + 0.018*“havaita“ + 0.017*“maanantai“ + 0.016*“välttyä“’),

90, ’0.071*“toivoa“ + 0.054*“sotilas“ + 0.048*“rakentaja“ + 0.044*“vaarallinen“ + 0.043*“hakata“ + 0.033*“niska“ + 0.032*“taistella“ + 0.029*“ilmaista“ + 0.028*“tammikuu“ + 0.028*“kutsu“ + 0.021*“visio“ + 0.021*“maito“ + 0.019*“tunnelma“ + 0.018*“piru“ + 0.018*“liittyminen“’),

91, ’0.060*“kieltää“ + 0.040*“luonto“ + 0.039*“sulkea“ + 0.033*“kunnioittaa“ + 0.025*“rakentua“ + 0.025*“purkaa“ + 0.022*“lopputulos“ + 0.020*“ympäristö“ + 0.018*“toistua“ + 0.018*“soveltaa“ + 0.017*“asettua“ + 0.016*“jättäminen“ + 0.014*“käytös“ + 0.014*“epäonnistua“ + 0.013*“käyttö“’),

92, ’0.092*“televisio“ + 0.054*“pienetä“ + 0.052*“kulutus“ + 0.051*“sarja“ + 0.040*“upea“ + 0.039*“aikaansaada“ + 0.031*“saati“ + 0.030*“painua“ + 0.022*“eväs“ + 0.022*“kituuttaa“ + 0.022*“lihava“ + 0.021*“naru“ + 0.018*“rajallinen“ + 0.017*“pässi“ + 0.016*“kovaa“’),

93, ’0.093*“suhde“ + 0.087*“ryhmä“ + 0.035*“yhdistys“ + 0.032*“tytär“ + 0.030*“syyllinen“ + 0.028*“ex“ + 0.025*“ylläpitää“ + 0.023*“arvio“ + 0.022*“levittää“ + 0.017*“toteuttaminen“ + 0.017*“useimmat“ + 0.016*“huolto#suhde“ + 0.016*“antaja“ + 0.016*“john“ + 0.014*“tarkoittaa“’),

94, ’0.060*“naapuri“ + 0.047*“puu“ + 0.039*“merkki“ + 0.034*“ovi“ + 0.028*“viro“ + 0.027*“korva“ + 0.027*“riita“ + 0.023*“rouva“ + 0.016*“pullo“ + 0.015*“väri“ + 0.014*“käydä“ + 0.014*“aita“ + 0.014*“kaivaa“ + 0.012*“japanilainen“ + 0.012*“maantie“’),

95, ’0.087*“järjestelmä“ + 0.081*“kaunis“ + 0.051*“sokea“ + 0.049*“varastaa“ + 0.046*“re“ + 0.040*“tuhota“ + 0.032*“palauttaa“ + 0.030*“perusoikeus“ + 0.030*“viestintä“ + 0.029*“kaunistaa“ + 0.025*“ihminen“ + 0.024*“suojelu“ + 0.023*“suvaitsevaisuus“ + 0.020*“ale“ + 0.016*“raitis“’),

96, ’0.093*“punainen“ + 0.075*“tavara“ + 0.065*“pöytä“ + 0.034*“puhemies“ + 0.030*“arvoisa“ + 0.030*“hieno“ + 0.029*“mon“ + 0.021*“pilkata“ + 0.016*“vahingoittaa“ + 0.016*“urakka“ + 0.015*“iltapäivä“ + 0.015*“ties“ + 0.014*“nk“ + 0.013*“piippo“ + 0.012*“muille“’),

97, ’0.231*“yrittäjä“ + 0.093*“vakuutusyhtiö“ + 0.069*“yritys“ + 0.052*“elättää“ + 0.025*“eläketurva“ + 0.024*“koettaa“ + 0.022*“pilata“ + 0.020*“ota“ + 0.019*“ihmetyttää“ + 0.018*“japani“ + 0.016*“yrittää“ + 0.016*“työterveyshuolto“ + 0.014*“työpanos“ + 0.014*“nolla“ + 0.012*“kori“’),

98, ’0.073*“eur“ + 0.049*“keskus“ + 0.047*“tuska“ + 0.040*“hätä“ + 0.038*“yksinhuoltaja“ + 0.033*“markkinointi“ + 0.023*“pussi“ + 0.022*“projekti“ + 0.022*“kuiva“ + 0.021*“sijaita“ + 0.021*“puola“ + 0.021*“puhdistaa“ + 0.020*“kutsua“ + 0.019*“oinas“ + 0.018*“äärimmäinen“’),

99, ’0.080*“hakea“ + 0.062*“myöntää“ + 0.057*“työkyvyttömyyseläke“ + 0.032*“kela“ + 0.030*“päätös“ + 0.024*“hakija“ + 0.024*“hakemus“ + 0.024*“työkyvyttömyys“ + 0.019*“kuntoutus“ + 0.019*“kuntoutus#tuki“ + 0.018*“määräaikainen“ + 0.016*“tieteellinen“ + 0.015*“perustaminen“ + 0.012*“työkyky“ + 0.012*“sii“’),

100, ’0.053*“tappaa“ + 0.040*“tottua“ + 0.040*“lopullisesti“ + 0.034*“pitkään“ + 0.033*“liika“ + 0.029*“pakollinen“ + 0.028*“äijä“ + 0.025*“haalia“ + 0.024*“jostain“ + 0.023*“pyörittää“ + 0.023*“pentu“ + 0.023*“vaivautua“ + 0.022*“älytön“ + 0.022*“pois“ + 0.018*“selvätä“’),

101, ’0.094*“firma“ + 0.053*“kouluttaa“ + 0.043*“nimittäin“ + 0.040*“sisältö“ + 0.035*“laulu“ + 0.030*“julkituoda“ + 0.028*“vallita“ + 0.022*“sairaanhoitaja“ + 0.020*“ohjaava“ + 0.020*“kerroin“ + 0.019*“puite“ + 0.018*“luottamus“ + 0.016*“avoimesti“ + 0.016*“loukata“ + 0.015*“ilmetä“’),

102, ’0.050*“silmä“ + 0.039*“valo“ + 0.027*“sydän“ + 0.026*“nähdä“ + 0.023*“alkaa“ + 0.020*“ruuma“ + 0.019*“valkoinen“ + 0.016*“tietää“ + 0.016*“ajatus“ + 0.015*“outo“ + 0.014*“kuluttua“ + 0.014*“kutsua“ + 0.013*“kadota“ + 0.013*“näyttää“ + 0.012*“tapahtua“’),

103, ’0.203*“julkinen“ + 0.123*“yksityinen“ + 0.063*“sektori“ + 0.043*“yhdistyä“ + 0.031*“palkankorotus“ + 0.027*“gallup“ + 0.015*“ottelu“ + 0.014*“international“ + 0.013*“moraalinen“ + 0.011*“kasvattaminen“ + 0.011*“kieroilu“ + 0.010*“alkupuoli“ + 0.009*“sittenkään“ + 0.009*“linja“ + 0.009*“malttaa“’),

104, ’0.144*“eläke#päivä“ + 0.135*“vääryys“ + 0.092*“viettää“ + 0.065*“nuoriso“ + 0.061*“onnettomuus“ + 0.034*“repiä“ + 0.024*“pääkaupunkiseutu“ + 0.022*“varmaankin“ + 0.021*“hyvinkin“ + 0.021*“puutarha“ + 0.020*“vähentäminen“ + 0.020*“kauempi“ + 0.019*“kerrostalo“ + 0.017*“ryypätä“ + 0.017*“luonne“’),

105, ’0.093*“päivä“ + 0.068*“ilta“ + 0.053*“aamu“ + 0.049*“yö“ + 0.036*“nukkua“ + 0.030*“alkaa“ + 0.028*“herätä“ + 0.023*“odotella“ + 0.021*“viikko“ + 0.020*“syksy“ + 0.017*“klo“ + 0.017*“uni“ + 0.016*“vuoro“ + 0.013*“ehtiä“ + 0.012*“syödä“’),

106, ’0.303*“yhteiskunta“ + 0.037*“kuluttaa“ + 0.026*“peruskoulu“ + 0.024*“lanka“ + 0.022*“rangaista“ + 0.019*“kko“ + 0.019*“myöten“ + 0.017*“pelkkä“ + 0.015*“asp“ + 0.015*“työllinen“ + 0.012*“rikko“ + 0.012*“sellanen“ + 0.012*“luovuttaa“ + 0.010*“parhaat“ + 0.010*“anarkia“’),

107, ’0.100*“vahva“ + 0.048*“rahoitus#tarkastus“ + 0.036*“vannoa“ + 0.034*“litra“ + 0.032*“laskeminen“ + 0.027*“realistinen“ + 0.025*“ilkeä“ + 0.021*“aluepolitiikka“ + 0.021*“kaksinkertaistua“ + 0.018*“etusija“ + 0.017*“vakuutuslaitos“ + 0.016*“lupailla“ + 0.015*“nolo“ + 0.014*“paljolti“ + 0.014*“virhi“’),

108, ’0.049*“toisten“ + 0.041*“isku“ + 0.035*“posti“ + 0.034*“energia“ + 0.028*“käyvä“ + 0.025*“työssäkäyvä“ + 0.025*“ihminen“ + 0.024*“vauva“ + 0.022*“usein“ + 0.021*“ankara“ + 0.021*“alko“ + 0.018*“leijona“ + 0.017*“ilman“ + 0.017*“sivistyä“ + 0.017*“kokonainen“’),

109, ’0.072*“palvella“ + 0.071*“todistaa“ + 0.055*“yk“ + 0.047*“voimakas“ + 0.038*“päämäärä“ + 0.032*“vaara“ + 0.031*“käyttää“ + 0.027*“pelastus“ + 0.026*“kappale“ + 0.019*“kisa“ + 0.019*“lakata“ + 0.018*“odotus“ + 0.017*“stä“ + 0.017*“termi“ + 0.014*“lopettaminen“’),

110, ’0.099*“ruumis“ + 0.067*“toimiva“ + 0.066*“teksti“ + 0.059*“dollari“ + 0.052*“ilmiö“ + 0.043*“inflaatio“ + 0.037*“aivot“ + 0.030*“käytävä“ + 0.028*“talli“ + 0.023*“välitys“ + 0.016*“kilpi“ + 0.015*“ihailla“ + 0.015*“napata“ + 0.013*“pahemmin“ + 0.012*“karpo“’),

111, ’0.089*“valhe“ + 0.058*“juhla“ + 0.050*“kannabis“ + 0.041*“sisäasiainministeriö“ + 0.036*“potku“ + 0.021*“hurja“ + 0.018*“ammattitaito“ + 0.018*“kärsimys“ + 0.017*“syrjä“ + 0.017*“järjestäytyä“ + 0.016*“uhkailla“ + 0.014*“pieli“ + 0.014*“kyseenalainen“ + 0.014*“salmi“ + 0.014*“psykiatrinen“’),

112, ’0.160*“puolue“ + 0.096*“keskusta“ + 0.048*“kepu“ + 0.043*“ehdokas“ + 0.040*“äänestää“ + 0.033*“vaalea“ + 0.028*“kokoomus“ + 0.027*“ääni“ +

0.024*“kannatus“ + 0.018*“kepulainen“ + 0.018*“perus-#suomalainen“ +
0.018*“poliittinen“ + 0.017*“kannattaja“ + 0.017*“äänestäjä“ +
0.016*“eduskuntavaalit“’),
113, ’0.141*“kansaneläke“ + 0.044*“muuttua“ + 0.029*“hymyillä“ +
0.026*“vanhuuseläke“ + 0.025*“voima“ + 0.022*“lukea“ + 0.018*“ks“ + 0.016*“liite“ +
0.016*“saksa“ + 0.015*“ts“ + 0.015*“laki“ + 0.014*“myöhä“ +
0.014*“elinkustannukset“ + 0.013*“henkilökohtainen“ + 0.013*“sairausvakuutus“’),
114, ’0.136*“yritys“ + 0.047*“suomi“ + 0.020*“miljardi“ + 0.017*“miljoona“ +
0.016*“poistaminen“ + 0.015*“teollisuus“ + 0.015*“elintarvike“ + 0.012*“tukea“ +
0.012*“tukiainen“ + 0.011*“maa“ + 0.010*“suomalainen“ + 0.010*“jakaa“ +
0.010*“ulkomaa“ + 0.009*“tukeminen“ + 0.009*“veroraha“’),
115, ’0.137*“jakaa“ + 0.049*“asiallinen“ + 0.044*“ilkka“ + 0.040*“lupa“ +
0.028*“laita“ + 0.023*“ärsyttää“ + 0.023*“kipeä“ + 0.022*“luuseri“ + 0.021*“kiinteä“ +
0.019*“erottaa“ + 0.019*“minuutti“ + 0.019*“kanerva“ + 0.018*“saatava“ +
0.018*“viel“ + 0.017*“julkisuus“’),
116, ’0.136*“eduskunta“ + 0.090*“kansanedustaja“ + 0.077*“edustaja“ +
0.045*“ministeri“ + 0.036*“etu“ + 0.031*“kausi“ + 0.023*“kansalainen“ +
0.016*“valita“ + 0.016*“päättää“ + 0.016*“tehdä“ + 0.015*“pitää“ + 0.015*“hedelmä“
+ 0.014*“palkka“ + 0.013*“nostaa“ + 0.011*“mitätön“’),
117, ’0.143*“omaisuus“ + 0.127*“arvo“ + 0.098*“myydä“ + 0.061*“omistaa“ +
0.035*“tappio“ + 0.032*“kateus“ + 0.026*“itsenäisyys“ + 0.025*“varallisuus“ +
0.019*“huolestua“ + 0.018*“tms“ + 0.013*“selitellä“ + 0.012*“tarkata“ + 0.010*“raha“
+ 0.010*“soitella“ + 0.010*“mukaa“’),
118, ’0.156*“aho“ + 0.055*“vasemmistolainen“ + 0.048*“ympäristöministeriö“ +
0.042*“halla“ + 0.040*“-jussi“ + 0.032*“pysäyttää“ + 0.029*“kulma“ + 0.023*“kärki“ +
0.021*“veikko“ + 0.018*“sorsa“ + 0.017*“merkittävä“ + 0.016*“kantola“ +
0.015*“vaalikampanja“ + 0.015*“metalli“ + 0.014*“kemia“’),
119, ’0.412*“vero“ + 0.053*“muista“ + 0.049*“lisä“ + 0.044*“poisto“ + 0.041*“kalle“
+ 0.036*“autovero“ + 0.029*“miinus“ + 0.029*“kuitata“ + 0.026*“maksaa“ +
0.024*“mieletön“ + 0.021*“bonus“ + 0.018*“kotitalous“ + 0.017*“sos“ +
0.017*“maksa“ + 0.015*“tuonti“’),
120, ’0.064*“kiva“ + 0.041*“vittu“ + 0.040*“huone“ + 0.039*“sairaseläke“ +
0.032*“saatana“ + 0.026*“hiljaa“ + 0.024*“siivota“ + 0.021*“lattia“ + 0.019*“työnteko“
+ 0.018*“vaikkei“ + 0.018*“vaik“ + 0.014*“levitä“ + 0.014*“yläkerta“ + 0.013*“kiinni“ +
0.012*“muija“’),
121, ’0.103*“markka“ + 0.102*“mk“ + 0.030*“opintotuki“ + 0.029*“milj“ +
0.020*“korkeakoulu“ + 0.018*“mrd“ + 0.017*“fortuma“ + 0.016*“kiinteistö“ +

0.015*“lintu“ + 0.014*“miljt“ + 0.012*“sisältyä“ + 0.012*“tehdä“ + 0.012*“määräraha“ + 0.011*“poiketa“ + 0.011*“koskea“’),
122, ’0.108*“vuotias“ + 0.097*“eläkeikä“ + 0.038*“täyttää“ + 0.036*“nostaa“ + 0.033*“päästä“ + 0.032*“ikä“ + 0.029*“jäädä“ + 0.027*“työ“ + 0.025*“ansiosidonnainen“ + 0.023*“päiväraha“ + 0.020*“iätä“ + 0.016*“työelämä“ + 0.015*“eläke#putki“ + 0.015*“pitää“ + 0.014*“syntyä“’),
123, ’0.049*“kela“ + 0.045*“summa“ + 0.032*“maksaa“ + 0.031*“raha“ + 0.026*“elää“ + 0.022*“ottaa“ + 0.020*“piirustus“ + 0.018*“pitää“ + 0.014*“kierros“ + 0.014*“tulo“ + 0.014*“kuukausi“ + 0.013*“pykälä“ + 0.012*“velka“ + 0.012*“tilanne“ + 0.011*“oikeuttaa“’),
124, ’0.037*“mielipide“ + 0.033*“tasavalta“ + 0.027*“sitoutua“ + 0.025*“virallinen“ + 0.024*“näkemys“ + 0.021*“politiikka“ + 0.021*“yhtyä“ + 0.021*“esittää“ + 0.019*“herättää“ + 0.019*“muodostaa“ + 0.016*“perinteinen“ + 0.015*“ehdoton“ + 0.014*“tärkeä“ + 0.014*“osoittaa“ + 0.013*“hesari“’),
125, ’0.064*“mitata“ + 0.050*“hlö“ + 0.041*“jono“ + 0.038*“muistutus“ + 0.025*“leipäjono“ + 0.022*“elävä“ + 0.020*“mätä“ + 0.019*“sääli“ + 0.017*“pidetä“ + 0.017*“pätkätyöläinen“ + 0.017*“puutteellinen“ + 0.016*“katkeruus“ + 0.016*“näinollen“ + 0.016*“piilo“ + 0.015*“sanoma#lehti“’),
126, ’0.076*“liitto“ + 0.019*“toimisto“ + 0.018*“työkkäri“ + 0.017*“työnhakija“ + 0.015*“pitää“ + 0.015*“kuulua“ + 0.014*“palautua“ + 0.014*“saaminen“ + 0.013*“syy“ + 0.012*“ilmoittautua“ + 0.012*“psykiatri“ + 0.011*“soveltua“ + 0.011*“ehto“ + 0.010*“työttömyys#turva#laki“ + 0.010*“työttömyysturva“’),
127, ’0.032*“pitää“ + 0.026*“tehdä“ + 0.018*“tietää“ + 0.017*“käydä“ + 0.016*“sanoa“ + 0.016*“alkaa“ + 0.014*“tuttu“ + 0.013*“ottaa“ + 0.012*“yrittää“ + 0.012*“juttu“ + 0.011*“päästä“ + 0.011*“osata“ + 0.011*“ihminen“ + 0.010*“mieli“ + 0.009*“paikka“’),
128, ’0.034*“toimittaja“ + 0.034*“rasismi“ + 0.033*“surkea“ + 0.031*“polkea“ + 0.029*“suuresti“ + 0.029*“kg“ + 0.027*“paino“ + 0.026*“kommunismi“ + 0.022*“sika“ + 0.021*“tyhmyys“ + 0.017*“epävarma“ + 0.016*“feminismi“ + 0.016*“asiaton“ + 0.015*“terrorismi“ + 0.015*“ylistys“’),
129, ’0.120*“maksu“ + 0.083*“maksa“ + 0.041*“leipä“ + 0.036*“sormi“ + 0.034*“pappa“ + 0.034*“eläkemaksu“ + 0.028*“perä“ + 0.026*“vanhainkoti“ + 0.025*“maksaa“ + 0.025*“puhdas“ + 0.015*“vapaaehtoinen“ + 0.014*“loppua“ + 0.013*“pieru“ + 0.012*“teho“ + 0.012*“eläke#vakuutus#maksu“’),
130, ’0.116*“työntekijä“ + 0.103*“pankki“ + 0.094*“työnantaja“ + 0.044*“maksaa“ + 0.040*“korko“ + 0.027*“palkka“ + 0.023*“ottaa“ + 0.017*“periä“ + 0.017*“laina“ +

0.016*“maksu“ + 0.014*“maksaja“ + 0.014*“irtisanominen“ + 0.013*“pitää“ +
0.010*“jäsenmaa“ + 0.009*“strategia“’),
131, ’0.059*“menestyä“ + 0.047*“lehtiä“ + 0.044*“työura“ + 0.034*“levy“ +
0.032*“varhaiseläke“ + 0.030*“varapuheenjohtaja“ + 0.027*“toimikausi“ +
0.026*“olemassaolo“ + 0.024*“langeta“ + 0.024*“ahdas“ + 0.019*“jakaminen“ +
0.018*“järkyttää“ + 0.018*“euroopata“ + 0.017*“rohkeus“ + 0.017*“vanhastaa“’),
132, ’0.081*“lista“ + 0.057*“kahvi“ + 0.056*“kala“ + 0.041*“ry“ + 0.041*“hiukka“ +
0.029*“ranta“ + 0.027*“sato“ + 0.026*“an“ + 0.024*“marja“ + 0.021*“kiertää“ +
0.019*“kiinnostus“ + 0.019*“itsenäisyyspäivä“ + 0.017*“tuijottaa“ + 0.016*“liha“ +
0.015*“tiitinen“’),
133, ’0.243*“poika“ + 0.048*“esitys“ + 0.044*“itkeä“ + 0.039*“testi“ +
0.036*“tehdas“ + 0.034*“halme“ + 0.028*“täti“ + 0.027*“perätä“ + 0.018*“kho“ +
0.015*“lasse“ + 0.015*“ulkoistaa“ + 0.014*“nyrkki“ + 0.014*“toiminimi“ +
0.014*“varkaus“ + 0.013*“millaista“’),
134, ’0.235*“seurakunta“ + 0.124*“kirkko“ + 0.048*“hengellinen“ +
0.035*“laamanni“ + 0.021*“tasku“ + 0.019*“kenraali“ + 0.019*“alkuseurakunta“ +
0.018*“luonnollinen“ + 0.015*“piispa“ + 0.013*“jäsen“ + 0.012*“koostua“ +
0.011*“kuopio“ + 0.010*“sananlasku“ + 0.010*“retki“ + 0.010*“ahkera“’),
135, ’0.080*“palsta“ + 0.056*“lukea“ + 0.045*“kirjoitus“ + 0.030*“selittää“ +
0.028*“eteenpäin“ + 0.026*“haista“ + 0.026*“lukija“ + 0.025*“foorumi“ +
0.021*“vieras“ + 0.021*“pahasti“ + 0.019*“perjantai“ + 0.019*“käydä“ +
0.018*“lukema“ + 0.017*“ketju“ + 0.017*“opetella“’),
136, ’0.148*“koulu“ + 0.126*“opiskelija“ + 0.100*“opettaja“ + 0.044*“huutaa“ +
0.038*“opettaa“ + 0.031*“oppilas“ + 0.026*“kiusata“ + 0.025*“luokka“ +
0.022*“rehtori“ + 0.019*“kiusaaminen“ + 0.016*“ahde“ + 0.013*“ainut“ +
0.012*“nyky“ + 0.012*“kasvo“ + 0.012*“maksumies“’),
137, ’0.432*“euro“ + 0.097*“miljoona“ + 0.043*“kuukausi“ + 0.021*“elää“ +
0.016*“laskelma“ + 0.015*“maksaa“ + 0.013*“kyl“ + 0.012*“tulo“ + 0.010*“liikevaihto“
+ 0.010*“sentti“ + 0.010*“raha“ + 0.007*“vuotaa“ + 0.007*“reilu“ + 0.006*“kulku“ +
0.006*“käyttää“’),
138, ’0.083*“tunnustaa“ + 0.064*“edustaa“ + 0.056*“viisas“ + 0.052*“uskovainen“ +
0.051*“kiittää“ + 0.045*“sanomat“ + 0.042*“siunata“ + 0.039*“säilyä“ +
0.033*“enemmistö“ + 0.030*“huonosti“ + 0.028*“kuuluisa“ + 0.025*“rehellinen“ +
0.024*“neitsyt“ + 0.024*“teos“ + 0.021*“jousimies“’),
139, ’0.195*“elvis“ + 0.030*“kuolema“ + 0.025*“kuvata“ + 0.016*“verno“ +
0.015*“kuolla“ + 0.014*“löytyä“ + 0.014*“presley“ + 0.014*“asianajaja“ +

0.013*“virallinen“ + 0.012*“arkku“ + 0.012*“nimi“ + 0.012*“lisa“ + 0.012*“väittää“ + 0.012*“aro“ + 0.010*“into“’),
140, ’0.067*“päättäjä“ + 0.065*“sukupolvi“ + 0.035*“kunnioittaminen“ + 0.032*“anne“ + 0.027*“nöyrä“ + 0.026*“omia“ + 0.025*“nykyaika“ + 0.020*“ritari“ + 0.019*“ruma“ + 0.019*“laji“ + 0.019*“vastaanottaa“ + 0.016*“lasta“ + 0.016*“tiedostaa“ + 0.016*“lyödä“ + 0.016*“turpa“’),
141, ’0.072*“eläke#ikä“ + 0.027*“nosto“ + 0.026*“kanava“ + 0.025*“nostaminen“ + 0.020*“eläkerahasto“ + 0.018*“nykyinen“ + 0.016*“nostaa“ + 0.016*“eläke#järjestelmä“ + 0.015*“vuode“ + 0.014*“pitkä“ + 0.011*“pitää“ + 0.010*“ikääntyminen“ + 0.009*“väestö“ + 0.009*“vara“ + 0.008*“tehdä“’),
142, ’0.064*“vanhus“ + 0.050*“hoitaa“ + 0.035*“hoito“ + 0.027*“hoitaja“ + 0.023*“pitää“ + 0.019*“tarvita“ + 0.013*“yhteiskunta“ + 0.013*“tehdä“ + 0.013*“ihminen“ + 0.011*“raha“ + 0.011*“sairaala“ + 0.010*“omainen“ + 0.010*“elää“ + 0.010*“koto“ + 0.010*“riittää“’),
143, ’0.217*“kk“ + 0.088*“tulo“ + 0.054*“asumistuki“ + 0.044*“toimeentulotuki“ + 0.026*“tuki“ + 0.020*“käsi“ + 0.020*“elää“ + 0.018*“luukku“ + 0.017*“rooma“ + 0.016*“jerusalemi“ + 0.013*“köyhyysraja“ + 0.012*“asua“ + 0.012*“yhde“ + 0.011*“näyte“ + 0.011*“vuokra“’),
144, ’0.077*“salata“ + 0.068*“hävitä“ + 0.055*“vaatimus“ + 0.047*“harkita“ + 0.040*“hyväkseen“ + 0.038*“poissa“ + 0.036*“kenkä“ + 0.029*“tämmöinen“ + 0.026*“käsittely“ + 0.022*“määräys“ + 0.021*“tekniikka“ + 0.021*“arkisto“ + 0.020*“vaatia“ + 0.020*“asettaa“ + 0.020*“perustuslakivaliokunta“’),
145, ’0.102*“nauttia“ + 0.066*“ulos“ + 0.055*“vahinko“ + 0.051*“maata“ + 0.046*“tekijä“ + 0.026*“rikollinen“ + 0.026*“touhu“ + 0.025*“välittää“ + 0.024*“tuho“ + 0.024*“käskeä“ + 0.021*“järjetön“ + 0.020*“hirvittää“ + 0.020*“päivittäinen“ + 0.019*“päätellä“ + 0.016*“nimike“’),
146, ’0.070*“laiton“ + 0.068*“kiinnostua“ + 0.049*“alennus“ + 0.044*“suhtautua“ + 0.035*“moraali“ + 0.027*“keskenään“ + 0.027*“kännykkä“ + 0.023*“otsikko“ + 0.022*“ettet“ + 0.021*“kysely“ + 0.020*“vastaaja“ + 0.019*“vähemmän“ + 0.017*“älykäs“ + 0.016*“yleistää“ + 0.015*“status“’),
147, ’0.073*“lähettää“ + 0.056*“asiakas“ + 0.052*“soittaa“ + 0.047*“sivu“ + 0.038*“paperi“ + 0.034*“tilata“ + 0.034*“netti“ + 0.028*“myyjä“ + 0.027*“kysellä“ + 0.027*“puhelin“ + 0.026*“ukko“ + 0.024*“tietää“ + 0.024*“tyyppi“ + 0.021*“mielenkiintoinen“ + 0.021*“sanoa“’),
148, ’0.252*“mies“ + 0.223*“nainen“ + 0.026*“tyttö“ + 0.015*“seksi“ + 0.014*“parisuhde“ + 0.014*“suhde“ + 0.011*“pitää“ + 0.010*“tietää“ + 0.010*“ero“ +

0.009*“mieli“ + 0.008*“ikäinen“ + 0.007*“puhua“ + 0.007*“pitkä“ + 0.007*“-vuotias“ + 0.007*“koto“’),
149, ’0.053*“osastopäällikkö“ + 0.051*“perustuslaki“ + 0.040*“keskinäinen“ + 0.036*“poliittinen“ + 0.031*“yhdysvallat“ + 0.025*“valtiovarainministeriö“ + 0.022*“kansallinen“ + 0.020*“vastainen“ + 0.018*“raportti“ + 0.013*“suomi“ + 0.012*“tasavalta“ + 0.011*“peruste“ + 0.011*“perätön“ + 0.011*“todeta“ + 0.010*“julkistaa“’),
150, ’0.087*“puoliso“ + 0.081*“mummo“ + 0.069*“puumerkki“ + 0.052*“avioliitto“ + 0.034*“naida“ + 0.032*“testamentti“ + 0.030*“kuolla“ + 0.026*“yhteinen“ + 0.024*“paine“ + 0.023*“leski“ + 0.017*“kota“ + 0.017*“kuolema“ + 0.016*“järjestys“ + 0.014*“avio“ + 0.014*“avoliitto“’),
151, ’0.056*“jaksaa“ + 0.053*“työ“ + 0.030*“tehdä“ + 0.026*“raskas“ + 0.020*“kurssi“ + 0.017*“päästä“ + 0.013*“pitää“ + 0.013*“pitkä“ + 0.011*“työelämä“ + 0.011*“toivo“ + 0.011*“sokeus“ + 0.011*“loppu“ + 0.010*“ikä“ + 0.010*“mieli“ + 0.010*“tilanne“’),
152, ’0.040*“korottaa“ + 0.032*“ostovoima“ + 0.031*“pienituloinen“ + 0.031*“korotus“ + 0.029*“sosiaaliturva“ + 0.029*“toimeentulo“ + 0.023*“elintaso“ + 0.022*“työ#markkina#tuki“ + 0.019*“köyhyys“ + 0.018*“veroton“ + 0.017*“perusturva“ + 0.017*“nostaa“ + 0.016*“perusosa“ + 0.015*“tulo“ + 0.015*“peruspäiväraha“’),
153, ’0.085*“sana“ + 0.085*“sanoa“ + 0.029*“raamattu“ + 0.027*“valtakunta“ + 0.026*“uskoa“ + 0.024*“sallia“ + 0.024*“synti“ + 0.022*“usko“ + 0.019*“erota“ + 0.018*“kristitty“ + 0.015*“tahtoa“ + 0.013*“valhe“ + 0.012*“kuulua“ + 0.012*“siunaus“ + 0.012*“perusta“’),
154, ’0.098*“paras“ + 0.086*“kuva“ + 0.074*“väkivalta“ + 0.071*“henkinen“ + 0.040*“kokea“ + 0.039*“asenne“ + 0.037*“kohdata“ + 0.021*“arvokas“ + 0.021*“tulkinta“ + 0.020*“ilmapiiri“ + 0.020*“koe“ + 0.020*“kokemus“ + 0.018*“mittari“ + 0.017*“seuraamus“ + 0.017*“henkisesti“’),
155, ’0.071*“piha“ + 0.029*“puhumattakaan“ + 0.029*“ylläpito“ + 0.028*“kampanja“ + 0.027*“kyllästyä“ + 0.027*“avaus“ + 0.026*“sittenkin“ + 0.025*“tapahtuma“ + 0.025*“halo“ + 0.024*“kritisoida“ + 0.023*“ulkona“ + 0.023*“yläpuoli“ + 0.023*“lauantai“ + 0.022*“parata“ + 0.021*“hymy“’),
156, ’0.083*“yms“ + 0.072*“kone“ + 0.071*“armeija“ + 0.071*“ohje“ + 0.068*“sopiva“ + 0.036*“pappi“ + 0.035*“lähtö“ + 0.031*“käsky“ + 0.024*“huolto“ + 0.023*“tyytyväinen“ + 0.021*“ennenkuin“ + 0.021*“löytyä“ + 0.020*“painaa“ + 0.015*“tarjous“ + 0.014*“linkki“’),

157, ’0.078*“vapaa“ + 0.074*“loma“ + 0.049*“kertyä“ + 0.046*“perustelu“ + 0.042*“sopia“ + 0.032*“päivätä“ + 0.031*“lyhyt“ + 0.030*“muutoksenhakulautakunta“ + 0.026*“jäätää“ + 0.023*“pitää“ + 0.020*“ohittaa“ + 0.017*“lääketieteellinen“ + 0.016*“erotus“ + 0.016*“viikko“ + 0.014*“kesäloma“’),
158, ’0.069*“keksiä“ + 0.040*“ruotsalainen“ + 0.039*“komentaja“ + 0.038*“paljastua“ + 0.031*“riisto“ + 0.030*“markkinatalous“ + 0.026*“kumota“ + 0.021*“aueta“ + 0.018*“kunnollinen“ + 0.016*“urheilija“ + 0.016*“riistää“ + 0.016*“syyllistää“ + 0.015*“arvosana“ + 0.015*“pystyä“ + 0.015*“otsa“’),
159, ’0.093*“poliisi“ + 0.038*“rikos“ + 0.032*“tutkia“ + 0.030*“laittomuus“ + 0.025*“syyttää“ + 0.021*“kertoa“ + 0.019*“epäillä“ + 0.019*“oikeus“ + 0.018*“poli“ + 0.018*“valistaa“ + 0.016*“ampua“ + 0.015*“tehdä“ + 0.015*“uhri“ + 0.014*“tapahtua“ + 0.013*“rikollinen“’),
160, ’0.222*“tieto“ + 0.171*“vastata“ + 0.127*“ilmoittaa“ + 0.049*“ilmoitus“ + 0.031*“osoite“ + 0.028*“kysymys“ + 0.021*“selvittää“ + 0.014*“kusettaa“ + 0.014*“akka“ + 0.013*“kohta“ + 0.012*“kt“ + 0.011*“verkko“ + 0.011*“lomake“ + 0.010*“istutus“ + 0.010*“oieta“’),
161, ’0.129*“alue“ + 0.082*“yhteistyö“ + 0.054*“historia“ + 0.035*“vauhti“ + 0.030*“ase“ + 0.024*“lukio“ + 0.023*“länsi“ + 0.021*“auki“ + 0.020*“kohtalo“ + 0.019*“huuto“ + 0.016*“paeta“ + 0.015*“aste“ + 0.015*“sanomalehti“ + 0.014*“johtopäätös“ + 0.012*“irak“’),
162, ’0.088*“tyhmä“ + 0.066*“tulos“ + 0.063*“väki“ + 0.055*“turku“ + 0.045*“fakta“ + 0.038*“pitäis“ + 0.035*“enempää“ + 0.033*“osaava“ + 0.024*“nenä“ + 0.021*“palkkio“ + 0.019*“sotkea“ + 0.019*“ahne“ + 0.019*“opintolaina“ + 0.016*“päinvastainen“ + 0.012*“voit“’),
163, ’0.141*“sak“ + 0.078*“asiantuntija“ + 0.063*“lama“ + 0.051*“aate“ + 0.039*“menestys“ + 0.036*“edelleen“ + 0.028*“tahtoa“ + 0.025*“pelkästään“ + 0.021*“piki#eläkeläinen“ + 0.019*“tuominen“ + 0.019*“setä“ + 0.018*“antikristus“ + 0.018*“korkki“ + 0.018*“aalto“ + 0.017*“lauri“’),
164, ’0.111*“raha“ + 0.061*“eläkevakuutus“ + 0.038*“säästää“ + 0.037*“peli“ + 0.037*“pelata“ + 0.029*“kortti“ + 0.028*“säästö“ + 0.024*“vakuutus“ + 0.019*“rahasto“ + 0.017*“huijata“ + 0.015*“käyttää“ + 0.014*“vapaaehtoinen“ + 0.013*“vara“ + 0.012*“vakuutusyhtiö“ + 0.011*“jälleenmyyjä“’),
165, ’0.040*“tutkimus“ + 0.025*“unioni“ + 0.021*“kehittää“ + 0.021*“kansalainen“ + 0.020*“kehitys“ + 0.017*“esittää“ + 0.017*“yhteiskunnallinen“ + 0.017*“kansainvälinen“ + 0.017*“kulttuuri“ + 0.015*“taso“ + 0.015*“lähde“ + 0.014*“osoittaa“ + 0.014*“sosiaalinen“ + 0.014*“synnyttää“ + 0.014*“pyrkiä“’),

166, ’0.082*“ehto“ + 0.053*“hallita“ + 0.051*“lupaus“ + 0.050*“linja“ + 0.036*“vammainen“ + 0.036*“vitsi“ + 0.031*“unohtua“ + 0.029*“polvi“ + 0.028*“huom“ + 0.027*“pekka“ + 0.027*“toisenlainen“ + 0.026*“mainos“ + 0.023*“valtuus“ + 0.022*“kyyti“ + 0.021*“sillo“’),
167, ’0.102*“aiheuttaa“ + 0.068*“alkoholi“ + 0.056*“sosiaali“ + 0.052*“käyttö“ + 0.047*“huume“ + 0.046*“terveys“ + 0.040*“terveysministeriö“ + 0.039*“kustannus“ + 0.030*“tauti“ + 0.022*“valvonta“ + 0.022*“tulkita“ + 0.018*“viina“ + 0.016*“oikeusministeriö“ + 0.015*“voimavarat“ + 0.014*“huumeisiin“’),
168, ’0.116*“jumala“ + 0.081*“jeesus“ + 0.057*“kristus“ + 0.046*“henki“ + 0.039*“totuus“ + 0.035*“pyhä“ + 0.028*“halko#saari“ + 0.026*“nimi“ + 0.018*“maailma“ + 0.017*“tie“ + 0.017*“ihminen“ + 0.016*“puhua“ + 0.014*“uskoa“ + 0.013*“palvelija“ + 0.013*“tehdä“’),
169, ’0.140*“laskea“ + 0.083*“palkka“ + 0.052*“työeläke“ + 0.037*“ansaita“ + 0.033*“määrä“ + 0.031*“vaikuttaa“ + 0.028*“tulo“ + 0.021*“paljonko“ + 0.018*“suuruus“ + 0.016*“työsuhde“ + 0.016*“huomio“ + 0.015*“elää“ + 0.015*“erikseen“ + 0.015*“ottaa“ + 0.014*“mestari“’),
170, ’0.073*“kuunnella“ + 0.059*“kaataa“ + 0.054*“pohjois“ + 0.054*“palestiinalainen“ + 0.041*“ira“ + 0.039*“nokia“ + 0.039*“karjala“ + 0.039*“palestiina“ + 0.029*“kannattaja“ + 0.026*“jäätteenmäki“ + 0.023*“brysseli“ + 0.022*“vinoutuma“ + 0.022*“savo“ + 0.019*“nooka“ + 0.018*“lainen“’),
171, ’0.036*“työttömyys“ + 0.030*“määrä“ + 0.025*“suomi“ + 0.024*“nykyinen“ + 0.017*“tarvita“ + 0.017*“kasvu“ + 0.017*“lisääntyä“ + 0.016*“kasvaa“ + 0.014*“ongelma“ + 0.014*“maahanmuuttaja“ + 0.013*“suomalainen“ + 0.012*“luku“ + 0.011*“tulevaisuus“ + 0.011*“maa“ + 0.011*“työvoima“’),
172, ’0.092*“poliitikko“ + 0.052*“kassa“ + 0.051*“jäsenyys“ + 0.031*“oiva“ + 0.029*“tentti“ + 0.022*“valvoa“ + 0.018*“ottaa“ + 0.018*“kansanvalta“ + 0.018*“kansanäänestys“ + 0.018*“raha“ + 0.017*“kohtelu“ + 0.016*“pahuus“ + 0.016*“torjua“ + 0.015*“piilottaa“ + 0.014*“omituinen“’),
173, ’0.091*“talo“ + 0.041*“asua“ + 0.035*“metsä“ + 0.031*“kesä“ + 0.025*“kilometri“ + 0.024*“km“ + 0.019*“sataa“ + 0.019*“isäntä“ + 0.019*“tontti“ + 0.017*“keskuuteen“ + 0.016*“hammas“ + 0.015*“tälläinen“ + 0.015*“talvi“ + 0.015*“etelä“ + 0.014*“keittiö“’),
174, ’0.122*“voittaa“ + 0.078*“näkökulma“ + 0.051*“tuleva“ + 0.047*“niin“ + 0.044*“myödä“ + 0.036*“lähi“ + 0.035*“jauhaa“ + 0.028*“hidas“ + 0.026*“tarkastella“ + 0.026*“korjaaminen“ + 0.025*“väkiluku“ + 0.024*“ajankohtainen“ + 0.020*“ravinto“ + 0.020*“apulaisjohtaja“ + 0.020*“kimmo“’),

175, ’0.080*“katsella“ + 0.039*“liikkua“ + 0.032*“sukulainen“ + 0.029*“kylmä“ + 0.026*“paikallinen“ + 0.023*“talvi“ + 0.019*“saari“ + 0.017*“pelottaa“ + 0.016*“jarmo“ + 0.016*“suosia“ + 0.015*“filosofia“ + 0.014*“taitaa“ + 0.014*“paikka“ + 0.014*“para“ + 0.013*“peto“’),
176, ’0.066*“määritellä“ + 0.045*“kello“ + 0.044*“etuus“ + 0.042*“lakkauttaa“ + 0.029*“välttää“ + 0.028*“automaattisesti“ + 0.028*“lakisääteinen“ + 0.026*“selkeästi“ + 0.023*“saanti“ + 0.023*“yksilöllinen“ + 0.021*“suorittaa“ + 0.020*“vihdoinkin“ + 0.016*“omaava“ + 0.016*“tehtävä“ + 0.015*“eläke#tulo“’),
177, ’0.091*“miljardi“ + 0.058*“laina“ + 0.049*“raha“ + 0.040*“kreikka“ + 0.035*“profeetallinen“ + 0.033*“pankki“ + 0.024*“imf“ + 0.023*“täydellisesti“ + 0.022*“ulosmitata“ + 0.021*“vakuus“ + 0.020*“talous“ + 0.019*“ajautua“ + 0.015*“konkurssi“ + 0.014*“myyminen“ + 0.013*“kaikenkaikka“’),
178, ’0.122*“jäsen“ + 0.064*“perustaa“ + 0.038*“sääntö“ + 0.037*“seuraaja“ + 0.035*“insinööri“ + 0.034*“toimia“ + 0.031*“laiva“ + 0.028*“huijaus“ + 0.026*“myynti“ + 0.022*“liiketoiminta“ + 0.018*“toiminta“ + 0.018*“säilyttäminen“ + 0.017*“relander“ + 0.017*“ote“ + 0.016*“kilpailija“’),
179, ’0.114*“prosentti“ + 0.052*“osuus“ + 0.043*“verrata“ + 0.031*“luku“ + 0.022*“korkea“ + 0.021*“muihin“ + 0.018*“alhainen“ + 0.017*“laskea“ + 0.016*“kohtuullinen“ + 0.015*“keskimääräinen“ + 0.015*“vastaavasti“ + 0.015*“vuosittain“ + 0.014*“taiteilija“ + 0.014*“budjetti“ + 0.013*“kasvaa“’),
180, ’0.053*“porvari“ + 0.046*“työläinen“ + 0.043*“kommunisti“ + 0.037*“sosialismi“ + 0.031*“kapitalisti“ + 0.030*“työväki“ + 0.027*“kapitalismi“ + 0.022*“hajaannus“ + 0.020*“mtv“ + 0.019*“sosialisti“ + 0.018*“yhdistää“ + 0.017*“propaganda“ + 0.017*“hitler“ + 0.015*“eriseuraisuus“ + 0.014*“vallankumous“’),
181, ’0.113*“oikeus“ + 0.103*“laki“ + 0.070*“sopimus“ + 0.048*“turvata“ + 0.022*“voima“ + 0.022*“säätää“ + 0.021*“perustua“ + 0.020*“artikla“ + 0.019*“ihmisoikeus“ + 0.019*“riippua“ + 0.019*“hyväksyä“ + 0.017*“lainsäädäntö“ + 0.014*“olennainen“ + 0.013*“alv“ + 0.012*“nykyinen“’),
182, ’0.159*“liike“ + 0.076*“ay“ + 0.074*“kantaa“ + 0.034*“lammas“ + 0.032*“yhtäläinen“ + 0.029*“ammattiliitto“ + 0.025*“pääasia“ + 0.025*“vähäosainen“ + 0.024*“kusi“ + 0.020*“pituus“ + 0.019*“paul“ + 0.019*“tuomas“ + 0.019*“ek“ + 0.019*“neljäs“ + 0.017*“viides“’),
183, ’0.082*“elämä“ + 0.070*“ihminen“ + 0.065*“elää“ + 0.019*“tehdä“ + 0.018*“ajatella“ + 0.017*“mieli“ + 0.016*“vanhempi“ + 0.016*“rakkaus“ + 0.013*“aikuinen“ + 0.012*“alkaa“ + 0.012*“perhe“ + 0.011*“ajatus“ + 0.011*“käydä“ + 0.009*“paha“ + 0.009*“välittää“’),

184, ’0.077*“vähäinen“ + 0.077*“väestö“ + 0.059*“tilasto“ + 0.032*“vertailu“ + 0.032*“työministeriö“ + 0.025*“suhteellinen“ + 0.024*“kasvava“ + 0.020*“kymmenys“ + 0.019*“kuulua“ + 0.018*“tilasto#keskus“ + 0.015*“ennenkaikke“ + 0.014*“aviopuoliso“ + 0.014*“tulotaso“ + 0.013*“vilkas“ + 0.013*“ev“’),
185, ’0.224*“tuki“ + 0.069*“lopettaa“ + 0.057*“maatalous“ + 0.040*“maanviljelijä“ + 0.037*“kustantaa“ + 0.028*“metsätalousministeriö“ + 0.025*“kansallinen“ + 0.024*“rikastua“ + 0.023*“vara“ + 0.021*“viljelijä“ + 0.018*“jako“ + 0.013*“maataloustuki“ + 0.013*“raha“ + 0.011*“veikata“ + 0.011*“veronmaksaja“’),
186, ’0.055*“professori“ + 0.053*“tohtori“ + 0.046*“tutkija“ + 0.034*“oikeuskansleri“ + 0.024*“onneton“ + 0.021*“rohkea“ + 0.021*“uudenmaa“ + 0.020*“putoilla“ + 0.019*“mauton“ + 0.018*“video“ + 0.018*“työnjohtaja“ + 0.018*“yliopisto“ + 0.017*“oikeusministeri“ + 0.016*“lakimuutos“ + 0.016*“ennestään“’),
187, ’0.101*“toiminta“ + 0.067*“järjestää“ + 0.043*“osallistua“ + 0.041*“uskonto“ + 0.036*“paikkakunta“ + 0.033*“järjestö“ + 0.032*“harjoittaa“ + 0.027*“terveydenhuolto“ + 0.025*“kristillinen“ + 0.016*“pasuuna“ + 0.014*“kyky“ + 0.012*“pidettävä“ + 0.012*“eritä“ + 0.012*“ohjaus“ + 0.012*“tanssia“’),
188, ’0.068*“keskustelu“ + 0.039*“kirjoittaa“ + 0.035*“aihe“ + 0.034*“joulu“ + 0.030*“viesti“ + 0.030*“keskustella“ + 0.029*“puolustusvoimat“ + 0.020*“kommentti“ + 0.019*“kirjoittaja“ + 0.019*“suomia“ + 0.018*“nimimerkki“ + 0.017*“kirjoitella“ + 0.016*“palsta“ + 0.014*“lukea“ + 0.013*“kirjoitus“’),
189, ’0.057*“pohja#osa“ + 0.054*“vakuutus“ + 0.039*“em“ + 0.037*“palkansaaja“ + 0.032*“julistus“ + 0.032*“yel“ + 0.021*“vakuuttaa“ + 0.020*“piiri“ + 0.018*“työtulo“ + 0.016*“muua“ + 0.016*“ruuan“ + 0.016*“äänestys“ + 0.014*“puoluesihteeri“ + 0.014*“maatalousyrittäjä“ + 0.013*“vaalitentti“’),
190, ’0.064*“eläköityä“ + 0.051*“pomo“ + 0.039*“kuljettaja“ + 0.036*“seutu“ + 0.036*“toive“ + 0.032*“virallisesti“ + 0.030*“lokakuu“ + 0.030*“elokuu“ + 0.025*“ryhti“ + 0.025*“vaellus“ + 0.024*“stt“ + 0.024*“sala“ + 0.021*“toteutua“ + 0.019*“haave“ + 0.019*“kuvio“’),
191, ’0.116*“tuote“ + 0.075*“ura“ + 0.054*“saavuttaa“ + 0.052*“arvostaa“ + 0.034*“nopea“ + 0.026*“yltää“ + 0.025*“menneisyys“ + 0.023*“harmi“ + 0.020*“liksa“ + 0.017*“näkyä“ + 0.016*“kaleva“ + 0.015*“poikkeuksellinen“ + 0.014*“koski“ + 0.011*“kyvyttömyys“ + 0.011*“kapi“’),
192, ’0.095*“kansalainen“ + 0.077*“nimi“ + 0.063*“tunne“ + 0.057*“uhri“ + 0.049*“lahja“ + 0.036*“ihminen“ + 0.032*“fiksu“ + 0.025*“tuntea“ + 0.024*“hävetä“ + 0.023*“leimata“ + 0.021*“piirre“ + 0.020*“huhu“ + 0.019*“ajattelu“ + 0.019*“lehmä“ + 0.016*“häpeä“’),

193, ’0.128*“uutiset“ + 0.119*“vaalia“ + 0.096*“keski“ + 0.060*“kotimaa“ + 0.055*“iltalehti“ + 0.042*“yle“ + 0.040*“aloite“ + 0.030*“yrittäjyys“ + 0.023*“vaalipiiri“ + 0.022*“kura“ + 0.019*“panostaa“ + 0.018*“ilta#sanoma“ + 0.018*“ddr“ + 0.016*“puheenvuoro“ + 0.016*“jyrkkä“’),

194, ’0.100*“lääkäri“ + 0.064*“potilas“ + 0.031*“lausunto“ + 0.028*“sairas“ + 0.023*“sairaus“ + 0.021*“sairasloma“ + 0.017*“työ“ + 0.016*“sairastaa“ + 0.016*“työkyvytön“ + 0.015*“tehdä“ + 0.015*“todeta“ + 0.014*“päästä“ + 0.014*“kela“ + 0.012*“kehoittaa“ + 0.012*“tilanne“’),

195, ’0.207*“ikäluokka“ + 0.034*“kuuluva“ + 0.030*“nalle“ + 0.028*“siivooja“ + 0.028*“somali“ + 0.028*“yks“ + 0.025*“siirtyä“ + 0.024*“pohjola“ + 0.021*“liietä“ + 0.020*“kutsumus“ + 0.017*“pummi“ + 0.016*“romani“ + 0.015*“häipyä“ + 0.014*“valta-#väestö“ + 0.014*“vedättää“’),

196, ’0.239*“kaupunki“ + 0.041*“hanko“ + 0.034*“hanke“ + 0.030*“elokuva“ + 0.029*“valtuuttaa“ + 0.028*“rakentaa“ + 0.022*“eri-#seura“ + 0.022*“liittäminen“ + 0.020*“lahti“ + 0.020*“koululainen“ + 0.014*“asukas“ + 0.013*“sijaita“ + 0.013*“salailu“ + 0.011*“lääkitä“ + 0.011*“muuttaa“’),

197, ’0.080*“neuvosto-#liitto“ + 0.051*“sotia“ + 0.047*“muslimi“ + 0.037*“vihollinen“ + 0.030*“tuomi#oja“ + 0.029*“jokapäiväinen“ + 0.028*“iraki“ + 0.023*“ii“ + 0.021*“läntinen“ + 0.020*“juhlia“ + 0.020*“europea“ + 0.019*“isänmaallinen“ + 0.019*“huuli“ + 0.018*“etsiminen“ + 0.016*“maailmansota“’),

198, ’0.057*“ala“ + 0.039*“pitää“ + 0.036*“halone“ + 0.030*“oppia“ + 0.025*“ihminen“ + 0.021*“opiskelu“ + 0.016*“puhua“ + 0.016*“tehdä“ + 0.014*“vanha“ + 0.013*“osata“ + 0.012*“tietää“ + 0.012*“ikä“ + 0.012*“alkaa“ + 0.011*“toisiaan“ + 0.010*“työ“’),

199, ’0.456*“lapsi“ + 0.109*“perhe“ + 0.030*“vanha“ + 0.012*“koti“ + 0.012*“huolenpito“ + 0.012*“ikäinen“ + 0.011*“toiset“ + 0.011*“huolehtia“ + 0.009*“iäkäs“ + 0.009*“ikä“ + 0.008*“isätön“ + 0.008*“vuotias“ + 0.008*“kasvatus“ + 0.008*“kasvaa“ + 0.007*“elatusapu“’)]

Appendix 4: Topic labels and words used in labelling them

Topic ID: Topic label	Words used in labelling the topic
0: Friends and hobbies	wife, pal, friend, to get to know, to go in for
2: President Sauli Niinistö and	foreign politics, [Finland's president Sauli] Niinistö, war, USA, NATO, media, Finland, an attack, Sauli
3: Getting by	to get along, to acquire, to do, money, to need, to live
4: Living in a place	residence, price, to buy, rent, to pay
6: Leadership	leader, head, government, chancellor, leadership
7: Medicine	medicine, suicide; to eat, to order, to use, to quit, to feed, background, according to, to drift; perfection, to hit, light, the size of, pass
9: Goodness	love, peace, perfect, fellow human being, security, soul, power of change; individual, to renew, home; warning, free/unfastened, betrayal, land, fatherland
10: Badness and prison	black, death, prison, to lie, to dare, bottom, castle (in Finnish slang, used of prison), prisoner; little, to fly, meter, safe, old age, surface, spiritualize
11: Finland and other nations	country, EU, Finland, world, Finland, Europe, Russia
13: Official argumentation	person, information, law, reason, party, right; [the thing] in question, action, accurate, to mean, according, this kind of, to require, to write down; labour office
15: Buying and selling used cars	car, motor, down side; to change, buy, owner, change, commercial, buyer; fault, km, renovation, h/km
20: Socio-economic groups	poor, rich, money, retiree, unemployed, poverty, student, family with children; National Coalition Party, Finland, people, to live, human, common; to hold
22: Entertainment	program, TV, culture, radio, approach
23: Political promises	Social Democratic party, member of the SDP, promise, to promise, Ben, National Coalition Party; to hold, to do, to cherish, to remember, to may, to raise, to seem, to forget; pensioner

25: Travelling	travel, place, to visit, hotel; week, hour, nice, couple, day, to get to, long, summer; to be found, pensioner, flag
26: State, big amounts and personnel	state, municipality, Finland; million, billion; personnel, hiring, to be paid, supplant, officer; Sonera (a big Finnish telecommunication firm), death certificate, to move, to take, main building
33: Welfare state economics	well-being, economy, workforce, welfare state, employment
35: Frustration over system called society	system, human, indifference, to be enough, arse; workplace, work, to employ, social security office, money, to have to, to get to, away, to take; to do
39: Society and development planning	human, society, world; consequence, reality, problem, future, to solve, to lead; to hold/like, on offer, to do , to live, to begin, mercy
40: Policy making	development, basic income, technology, foreign policy, industry; active, to ensure, according to a view, adviser, driving, actor; to cut, to break, fool, meaning
41: Voting and different parties	to vote, green, labourer, defender, left wing
43: Eating	mouth, food, water, to sit, full
45: Happiness	to be satisfied, joy, happy; home, sun, life, scene, eternal, freedom; step, around, piece, me, to raise, to spend
50: Government	general director, [ex-prime minister Matti] Vanhanen, Matti, government, prime minister
51: Judging	to judge, judgement, layer, judge, act
53: Presidents	president, [president Tarja] Halonen, Tarja, Sauli, [president Mauno] Koivisto, [president Martti] Ahtisaari, Bush; Minister for Foreign Affairs, righteous, to choose; godless, book, message, comrade, qukoa (not a word)
54: Co-operation negotiations	to grow, negotiation, to lay off, employee, co-operation negotiations
55: Family, birth and death	mother, father, brother; to die, birth, old; sad, to be disappointed; he who brings, truly, along, building, music, effort, to warn
57: Gates in handling an application from a person's perspective	to wait, to let, hurry, entrance, to get in, extension positive, retirement decision, actively, vihta (a traditional Finnish massage equipment), voice, to choose, fear, poor, fresh

58: Working, time and salary	month, full, to earn, may, working time, day, salary, waste, week, to live
62: Payments and compensations	to pay, compensation, pay (imperative), money, paying, to compensate, cost, salary, tax funds, expensive, pension payment; tax payer, state, size of; accident
63: Treating illness	treatment, illness, depression, problem, ailment
68: Policy making	service, security, renewal, theme, day care, social, resources side, Luoma, important, partly, to limit, solution, common sugar
69: Studying for a profession	education, profession, to study, working life, degree
72: Government's cuts	government, to cut, cut, to do, children allowance
74: Age, retirement and activities	young, old, retiree/pensioner, wrinkle; to get into, to do, to stay, work, to begin, gang, to know how, to have to, to run to be saved, best
75: Names	amp (not a word), Juha, Erkki, Seppo, Liisa
78: Investing	profit, revenue, firm, share, to invest
79: Politics and policy makers	people, [famous politician Paavo] Lipponen, Finland, Paavo, politics, [famous politician Sinikka] Mönkäre, [former president Arto] Paasikivi
80: Work and livelihood	work, unemployed, salary, to hire, task, employer, work place, paid work; to need, to live, unemployment compensation, money; to do, to get to, day
83: Taxation	taxation, tax, to remove, just, income
84: Finland	Finland, Finnish, country, Finland/to whip, foreign country
86: Leaders, banks and insurance companies	Jukka, Sailas, CEO, Harri, Risto, Lilius, Mikael, Sampo-bank, Varma-insurance company, Ilmarinen-insrance company, option (to buy), reward, apprehension, top, witness (all names are thos of famous CEO's or such)
99: Applying and the benefits for the disabled	to apply for, to allow, decision, applicant, application, disability retirement, Kela (The Social Insurance Institution of Finland), disability unemployment, rehabilitation, rehability allowance, working ability, temporary, scientific, basing, sii (not a word)

105: Time and sleeping	day, evening, morning, night, week, autumn, at (of time) to sleep, to wake up, sleep, to wait, to begin, to be on time, to eat, shift
112: Voting for the Center Party	party, Center Party, Kepu (a nick name for the party), candidate, to vote
116: Members of parliament	parliament, member of parliament, deputy, minister citizen, to choose, to decide, to do, to hold, salary, to rise, benefit, season fruit, meaningless
117: Property	property, value, to sell, to own, defeat
118: Politicians	Aho, leftist, Ministry of the Environment, Halla[-Aho], Jussi, to stop, corner, head, Veikko, Sorsa
122: Getting older and out of working life	of age, retirement age, to turn, age, to get older, to be born; to rise, work, working life; earnings-dependant, daily benefit, retirement pipe; to hold, to get, to stay
123: Money	Kela, amount, to pay, money, income, month, loan, to live, to take, to hold, clause, situation, to justify drawing, round,
126: Employment and unemployment	employment exchange office, employee, act of protection from unemployment, protection from unemployment, union, office (place), to be part of, getting, reason, to enrol in, to be suited for, condition, psychiatric; to consider, to recover
127: Doing things	to have to, to do, to know, to go to, to say, to begin, to take, to try, to get in, to know to, human, mind, place, thing, familiar
129: Payments and pension	payment, pay (imperative of to pay), pension payment, to pay, threshold, pension insurance payment elderly home, grandpa; finger, bread, back, clean, voluntary, to finish, fart, effect
130: Payments and working life	bank, to pay, threshold, to inherit, loan, payment, payer employee, employer, salary, resigning to take, to hold, member state, strategy
134: Religious institutions	congregation, church, spiritual, lawspeaker, first congregation and bishop member, to be built of, excursion, diligent pocket, general, natural, Kuopio, proverb
135: Commenting writings in the forum	column/forum, to read, writing, to explain, forward
136: Schools	school, student, teacher, to shout, to teach
137: Money	euro, million, to pay/to cost, income, revenue, cent, money, month, to live, calculation, to use to leak, fair, path, for sure

138: Ecclesiastic activities	to confess, to represent, wise, believer, to thank
139: Elvis is dead	Elvis, Presley, death, to die, coffin
141: Pension system	retirement age, pension fund, pension system; increase, increasing, current, to rise, ageing, population; bed, long, to hold, spare, channel, to do
142: Elderly care	old person, to nurse, treatment, nurse, to keep
143: Social benefits	month, income, housing allowance, income support, benefit
147: Managing daily tasks via different channels	to send, to call, to order, page, paper, internet, phone client, salesman to ask, guy, to know, dude, interesting, to say
148: Intimate relationships	man, woman, girl, sex, intimate relationship,
149: Politics	head of department, constitution, central, political, United States of America
150: Marriage and death	spouse, marriage, to marry, testament, to die, common, widow, death, marital, common-law marriage granny, insignia, pressure, goahti (not a word), order
151: Hard work/ work is hard	to have the energy for, hard, long, end, work, to do, course, working life, age, blindness, to get to, to hold, hope, mind, situation
152: Basic social income support	to raise, purchasing power, person with low income, increase, social security, livelihood, standard of living, labour market support, poverty, tax-free, basic security, to raise, basic part, income, basic daily support
153: Religious beliefs	word, to say, the Bible, kingdom, to believe
154: Interpretations of experiences	to experience, attitude, to face, interpretation, atmosphere, experience, mentally, mental, best, picture, violence, valuable, exam, sensor, consequence
158: History	to invent, to reveal, deprivation, market economy, to abolish, to deprive, Swedish, commander, to blame, sportsman, forehead, grade, proper, to open, to be capable of
159: Crimes	police, crime, to investigate, illegality, to accuse
160: Announcements	information, to reply, to announce, announcement, address

161: War	area, cooperation, history, weapon, west, destiny, to escape, newspaper, Iraq, scream speed, high school, open, degree, conclusion
163: SAK	SAK (i.e. Central organisation of Finnish Trade Unions), expert, recession, ideology, success
164 Household savings and games	money, pension insurance, to save, saving, insurance, fund, insurance company; game, to play, card, to cheat, to use, voluntary, spear, reseller
165: Social development	union, to develop, citizen, development, societal, international, culture, social; research, to present, to show, source, level, to give birth, to aim
167: Drugs and society	alcohol, drug, illness, health, spirit, into drugs to cause, usage, cost, surveillance; energy resources, social, Ministry of health, Ministry of Justice, to interpret
168: Christianity	god, Jesus, Christ, spirit, truth
169: Salary	to decrease, salary, employee's pension, to earn, amount
170: Political places	northern, Palestinian, Nokia, Karjala, Palestine, Bryssel, Savo, "lainen" = attribute referring someone being from somewhere, as n in Macedonian supporter, bias, to listen, to fall, Jäätteenmäki, nooka*, IRA
171: Development in the Finnish society	unemployment, labour force; increase, to increase, to grow, amount, number; problem, future, country, Finland, current; immigrant, Finnish; to need
173: Living in a place and nature	house, to live, lot (as in real estate), kitchen forest, summer, to rain, wintern, south, kilometre, km, master, midst, tooth, this kind of
176: Benefits	benefit, pension income, statutory, automatically, individual, to suspend, to avoid, to perform, possessing, task, to define, supply, watch, clearly, finally
177: International loans	billion, loan, money, bank, IMF, bankruptcy, selling, warranty, economy, to repossess Greece, prophetic, perfectly, to drift, all in all
178: Business and associations	member, to found, rule, follower, to act, action business, competitor, selling, engineer, ship, maintenance; hoax, Relander, hold
179: Comparing quantitatively	percentage, share, to compare, number, high
180: Social classes and ideals	merchant, labourer, communist, socialism, capitalist
181: Legislation	justice, law, contract, article, human right, legislation to secure, to adjust, to base on, to depend on, to accept, essential, vat, current force

183: Human life	life, human, to live, to do, to think
184: Population statistics	minor, population, statistics, comparison, Ministry of Employment
185: Agricultural subsidy	farming, farmer, Ministry of Forestry, planter, support, to sponsor, Agricultural subsidy, share, money, tax payer to get rich, spare to quit, national, to guess
186: Designations	professor, doctorate, researcher, attorney general, foreman university, Ministry of Justice, miserable, brave, distasteful; Uusimaa, to fall, video, change in law, previously
187: Activities	action, to organise, to participate, organisation, to practice, religion, municipality district, health care, Christian, bassoon, ability, amiable, guidance, to dance, to differ
188: Forum posts	discussion, to write, topic, message, to discuss, comment, writer, nickname, to write, forum, to read, writing; Christmas, Finland, defence force
192: Stigmatisation	citizen, name, feeling, victim, gift, human, smart, to feel, to be ashamed, to stigmatize, feature, gossip, thinking, cow, shame
193: news	news, Iltalehti-magazine, YLE (the Finnish Broadcasting Company), Iltasanomat; initiative, to cherish (or a problem of the lemmatizer, must be rather related to elections (vaalit) than to cherish (vaalia)), say, electoral district; entrepreneurship, invest; middle, dirt, steep, homeland, DDR
194: To be ill	doctor, patient, certificate, ill, illness, sick leave, work, to be ill, working disabled; to do, to state, to get in to, KELA, to suggest, situation
196: Municipalities	city, Hanko, Lahti, conjunction, citizen, to be located; to move, to build; project, film; to authorise, ”different company”, school girl/boy, hiding, to medicate
197: Nations, cultures and war	Soviet Union, Muslim, Tuomioja, Iraq, western, Europe, patriotic; to be at war, enemy, world war, II; daily, to party, lip, search
199: Taking care of children and the elderly	child, family, old, home, care taking, of age, others, to take care of, elderly, age, fatherless, years old, upbringing, to grow, alimony

Appendix 5: Lists of topics in retirement related and general discussion used for comparing different and overlapping themes.

Retirement related topics	Corresponds with a topic in general discussion, which?	Theme
194: To be ill	y, 3, 33	Illnesses
63: Treating illness	y, 3, 33	Illnesses
7: Medicine	y, 91, 108	Illnesses
150: Marriage and death	n	Relationships and Family
55: Family, birth and death	n	Relationships and Family
173: Living in a place and nature	n	Living in a place
154: Interpretations of experiences	n	Life and Feelings
183: Human life	n	Life and Feelings
45: Happiness	n	Life and Feelings
9: Goodness	n	Life and Feelings
4: Living in a place	y, 148	Living in a place
148: Intimate relationships	y, 156	Relationships and Family
160: Announcements	n	Media
22: Entertainment	n	Media
193: news	y, 151	Media
135: commenting writings in the forum	y, 109	meta
188: Forum posts	y, 109, 112	meta
164: Household savings and games	n	Money
177: International loans	n	Money
178: Business and associations	n	Money
86: Leaders, banks and insurance companies	n	Money
117: Property	n	Money
123: Money	n	Money
137: Money	n	Money
15: Buying and selling used cars	n	Money
78: Investing	y, 71	Money

112: Voting for the Center party	n	Politics
116: Members of parliament	n	Politics
118: politicians	n	Politics
163: SAK	n	Politics
170: Political places	n	Politics
186: Designations	n	Politics
196: Municipalities	n	Politics
23: Political promises	n	Politics
26: State, big amounts and personnel	n	Politics
50: Government	n	Politics
53: Presidents	n	Politics
72: Governement's cuts	n	Politics/Social security system
79: politics and policy makers	n	Politics
83: Taxation	n	Politics/Social security system
158: History	n	Politics
33: Welfare state economics	y, 135	Politics
11: Finland and other nations	y, 16, 55, 134	Politics
161: War	y, 173	Politics
41: Voting and different parties	y, 20	Politics
197: Nations, cultures and war	n	Politics (international)
2: President Sauli Niinistö and foreign politics	n	Politics (international)
149: Politics	n	Politics/Social development
68: Policy making	n	Politics/Social development
40: Policy making	n	Politics/Social development
181: Legislation	y, 87	Politics/Social development
13: Official argumentation	n	Regulation
147: Managing daily tasks via different channels	n	Regulation
179: Comparing quantitavely	n	Regulation
134: Religious institutions	n	Religion

138: Ecclesiastic activities	n	Religion
153: Religious beliefs	y, 119	Religion
168: Christianity	y, 6, 130, 144	Religion
122: Getting older and out of working life	n	Retirement transition
74: Age, retirement and activities	n	Retirement transition
165: Social development	n	Social development
171: Development in the Finnish society	n	Social development
180: Social classes and ideals	n	Social development
35: Frustration over system called society	n	Social development
39: Society and development planning	n	Social development
20: Socio-economic groups	n	Social issues
10: Badness and prison	n	Social issues
167: Drugs and society	n	Social issues
192: Stigmatisation	n	Social issues
184: Population statistics	n	Social issues
159: Crimes	y, 114	Social issues
136: Schools	n	Social security system
141: Pension system	n	Social security system
142: Elderly care	n	Social security system
199: Taking care of children and the elderly	n	Social security system
143: Social benefits	n	Social security system
152: Basic social income support	n	Social security system
176: Benefits	n	Social security system
99: Applying and the benefits for the disabled	n	Social security system
57: Gates in handling an application from a person's perspective	y, 1	Social security system
0: Friends and hobbies	n	Spare time
105: Time and sleeping	n	Spare time
127: Doing things	y, 44	Spare time
187: Activities	y, 44	Spare time
25: Travelling	y, 53	Spare time
43: Eating	y, 74	Spare time
126: Employment and unemployment	n	Work and livelihood
130: Payments and working life	n	Work and livelihood
151: Hard work/ work is hard	n	Work and livelihood

169: Salary	n	Work and livelihood
3: Getting by	n	Work and livelihood
54: Co-operation negotiations	n	Work and livelihood
58: Working, time and salary	n	Work and livelihood
62: Payments and compensations	n	Work and livelihood
129: Payments and pension	n	Work and livelihood
185: Agricultural subsidy	n	Work and livelihood
69: Studying for a profession	y, 177	Work and livelihood
80: Work and livelihood	y, 76	Work and livelihood
139: Elvis is dead	n	
6: Leadership	n	
75: names	n	
84: Finland	n	
51: judging	y, 61	

All messages topics	**Corresponds with a topic in retirement discussion?**	**Theme**
122: body weight	n	body and illness
58: Fit	n	body and illness
59: Physical features	n	body and illness
170: therapy, depression	n	body and illness
18: Variables of anatomy	n	body and illness
108: Medical treatment	y	body and illness
3: Ailments and tips for treatment	y	body and illness
33: Ailments, treatment, effects	y	body and illness
91: Doctors and medicine	y	body and illness
57: Violence crimes	n	crimes
114: Crimes	y	crimes
89: Living in a house/renting	n	house and living in a place
73: To build a house	n	house and living in a place
98: Renovating	n	house and living in a place

110: Weather and places	n	house and living in a place
117: Features of houses	n	house and living in a place
145: places	n	house and living in a place
22: Seasons	n	weather
99: Temperature	n	weather
148: living in a place	y	house and living in a place
158: political news	n	media
17: Headline posters	n	media
52: Websites	n	media
66: Memes	n	media
12: News	y	media
151: news	y	media
113: Discussions	n	meta
109: Forum posts	y	meta
112: Posts	y	meta
14: Loans	n	money
142: commerce	n	money
90: Buying and selling	n	money
71: Trading and calculations	y	money
140: Finland's history and Germany	n	politics
16: Finninsh and Swedish	n	politics
187: fighting and terrorism	n	politics
2: Finland and immigrants	n	politics
25: Municipalities' government	n	politics
37: Russia and military	n	politics
43: Center party	n	politics
49: State economics	n	politics
56: True Finns	n	politics
134: Foreign relations	y	politics
135: Welfare state economics	y	politics
173: war tactics	y	politics

20: Voting and parties	y	politics
55: Countries and politics	y	politics
1: Boureaucracy	y	regualtion
61: Judgement	y	regualtion
87: law, regulations	y	regualtion
111: Relationships & cheating	n	Relationships, family
139: feelings of love	n	Relationships, family
162: money, relationship and sharing	n	relationships, family
164: growing up and family	n	Relationships, family
41: Boys and girls	n	relationships, family
75: Menstruation and pregnancy	n	Relationships, family
171: family	n	Relationships, family
77: Family living	n	Relationships, family
93: Sociality	n	Relationships, family
156: intimate relationship	y	Relationships, family
106: Religion	n	religion
28: Religion and science	n	religion
119: Religious beliefs	y	religion
130: Christianity	y	religion
144: Christianity	y	religion
6: Christianity	y	religion
175: reading	n	Spare time
31: Games & playing	n	spare time
4: Hunting	n	Spare time
46: Sleeping and waking up	n	spare time
79: Drinking booze	n	Spare time
126: Dog training	n	Spare time
160: photography	n	spare time
82: Dogs	n	Spare time
44: Spare time activities	y	spare time
53: Travelling	y	spare time
74: Meals	y	spare time
107: Computers and fixing computer problems	n	Technology
149: Phones and fixing and buying them	n	Technology

161: digital messages	n	Technology
197: devices and listening to music	n	Technology
27: Broken technology	n	Technology
29: Computer technology	n	Technology
36: Phone calls and contact information	n	Technology
96: Computers, OSs, software	n	Technology
174: motor vehicles	n	Transportation, vehicles
190: motor vehicles	n	Transportation, vehicles
62: Cars' features	n	Transportation, vehicles
11: Public transportation	n	Transportation, vehicles
48: Traffic and accidents	n	Transportation, vehicles
8: Traffic and features of vehicles	n	Transportation, vehicles
183: time and money	n	work and livelihood
185: income support decisions	n	work and livelihood
42: Entrepreneurship	n	work and livelihood
76: Work and livelihood	y	work and livelihood
177: getting to studies and work	y	work and livelihood
0: Human, smart	n	
167: swearing	n	
65: Services/ customer service	n	
78: Science	n	

Appendix 6: Conversations selected for the analysis

Of the topic number 122, the conversations number 1221, 11448, 5977, 10806, 31, 3415, 4954, 7315, 7527, 10039, 3067, 7369, 537 and 6401 were selected, and from topic number 74, the conversations number 1993, 3478, 8153, 5537, 329, 863, 7875, 2762, 8006, 6364, 3352, 8404, 6271 and 7584.

31, Pension benefit changes

The discussion of Pension benefit changes constituted of four messages. The first writer was bitter about the cuts of different social security benefits, including pensions, contrasted to the simultaneous increases in the agricultural subsidy. The second message was largely not related to retirement, commenting the different politicians, stating that their irresponsible political decisions have led to the current, miserable economic condition of Finland, and, the reason why the message got through the selection, declaring the writer's concern over "the unemployed, the poor and the pensioners". The last two messages may be translated as "Aren't the parasites of the social security system comforted for the good times foreseen for the pensioners?" and "Everyone's pension has to come down. 2000 euros will do."

537, Pensioners who work

The second message was full of relevant content. The initial writer pondered over the working pensioners in a municipality. Their opinion was, their places should be given to the young, competent workers. The second gave a reason for why they still were working, that it was because "is it not on everyone's lips that the retirement age should be increased otherwise the national economics will not be rebuilt." The third's explanation does not seem to appreciate the older workers, either, as they claim that the workers there are used to receive their pay check for barely hanging on their work places, so they don't have the patience to stay away even on retirement. The last writer in the discussion tree criticised that those on retirement and raising pension should not be allowed to raise the salary too, as the money would be much more needed in the hands of a younger person.

7, Pension pipe

The discussion I labelled Pension pipe consists of three messages, a question and two answers. The question is about how the possible layoff will affect pension accumulation in different scenarios. The answer, with the nickname of "Old age pension sage" is detailed and gives references.

Similar, question - answer -discussions as that held in the Pension pipe were the discussions number 5977, 7527 and 10806, respectively called At the margins of pension renewals (wondering how different her pension accumulation would have been had the current law been in act since her youth), Which pension application to fill in (where the questioner was fighting through the bureaucracy to receive pension) and Pension calculated correctly (where the writer asked how they could make sure that their pension was calculated correctly).

3067, Three things to be thankful for

The messages in the tree discuss, as name implies, the things that are making the commentators happy. There have likely been much more messages in the discussion tree, but with the restriction to the retirement word inclusion, there were only four left. Why retirement is mentioned is because one of the writer expects to have an amiable job hopefully until retirement, because the second writer appreciates her free time that retirement has provided her with a lot, because the third thinks that those three things are “bank account, pension [and] extra pension”, and because the fourth says that “none of the things have nothing to do with my pension being small or big“.

3415, Pension benefit vs. cost of prisoner

There was only one message in the discussion. It compares the cost of a life sentence to that of pension benefits. Her opinion was, it is the wrong priority order that pensions are not bigger but prisoners are given over 5000 euros monthly.

4954, Lowering retirement age

Retirement age was told in the discussion to be going to be lifted because there are worries that pension funds will otherwise dry out. However, the writer said, the funds are still growing so the renewal would only be bullying those who have already had their share of work. Besides, the writer preaches, the older workers who are keeping up in their work are blocking the places from the young, and those of them who cannot work are eating up the unemployment benefits. They advise, “to retirement those who want and jobs to the young, that will work for the benefit of our country”.

6401, Trade unions are fooling us

As the name implies, the writers in the two messages share strong opinions about trade unions. Somehow the pensions of the Greek ended up in the first message, and the second reminds that everyone, including pensioners, pay taxes. Unfortunately for the thesis’ analysis, the messages included the words *to fill, to stay, work, salary-bind* and

daily allowance that were members of the fifteen most probable words in the topic number 122, and thus it was available for the carefully constructed script that randomly selected the messages for the analysis.

And the *Trade unions...* was not the only badly selected discussion tree within the sample. There were also the messages indexed 7315, 5537, 6271 and 8404 which were called **Farmers as parasites** (for it held animosity towards farmers), **Immigration** (that went through the problematisation of the party True Finns concerning the issue), **Irresponsible entrepreneurship** (that went on to despise an irresponsible entrepreneur) and **Work is worth because of pension** (as that was given as a reason why one should work in the first place). According to the writers in the first of these discussions, farmers are living on the others, retire already in their 50's and receive big pensions for nothing. Despite the poor level of correctness, the discussion did concern retirement. The beef in the second discussion for this thesis was that it mentioned pensioners can also be immigrants. The third discussion worried that there are entrepreneurs who do not care for their pension security, and the fourth corrected the idea of someone else that it is not worth working with a salary less than 1000-2000 euros.

7369, Layoff is not a layoff after retirement age

This lone-standing message claimed, once you have reached your retirement age it is not possible to talk about laying off, it is only natural move to retirement.

The message chain indexed **1993, The young will never get a job if the pensioners will keep on working**, is pretty much explained by its label. The writer firmly suggests that pensioners should not work, and went even so far as to say, they "should be fined if they did go working once they have received pension".

10039, Retirement age lowering as a solution to unemployment?

One more chain of messages debated retirement age and work during retirement. The first writer thought, keeping retirement age as it is and denying work while receiving pension would be a great solution to decreasing unemployment. Another writer claimed, retirement age could be lifted as high as ever, but there are few of those who remain able to work after 65 years of age.

The discussion indexed **3478, Lifting the retirement age as a solution to dependency ratio,** was giving another political advice on how to handle state economics. The only message is of good quality, supporting the older workers and criticising the employers who won't let them stay. Employers are dismissed for their lack of realising the potential of older workers. Instead, they are hiring foreigners because they are younger and, as the employers except, have lesser expectations about salaries - although they

did not even ask what the older Finns would like to receive. The writer herself would be ready to work with a smaller salary, and wished the government to organise a wide survey of what salary would suffice citizens and how long a career would they really like to have.

11448, The rich

Ii was hard to confirm if the two messages in this conversation were genuine in their claims about their fortunes or if they were just barking. It was decided to take them seriously; in any case, no word in the forum can be confirmed to be what a person really thinks as it is very common to troll in the forum. However, the contents of the conversations were the other writer claiming they are only 36 years old and already retired with huge possessions, pursuing their dreams. The other, also very rich, gave tips to the forum visitors how to shorten the time one has to pay the mortgage.

329, Re-education

A short message of only four phrases received a lot of attention from me, as it said, “I cannot continue in my current work until retirement, so I have to start preparing for changing my profession”. More on the meanings that I gave to the message will follow in the analysis.

863, Quitting driving taxi

Another one-writer discussion, where a taxi driver analysed the condition of the field to be so bad that they would quit soon.

2762, The dependants

If there was a stereotype of the quality of a discussion tree on the forum before beginning to write the thesis, it would probably have been thought to be something like the messages on the dependants of the society. Argumentation was not based on facts but on ideas that were heard somewhere and on so-called “cold facts”. From pensions the writers jump to moralise all of the condition of the society. To quote the last writer: “... What I personally know from the last 10 years, the young women are making babies as lone parents and make the state pay alimony. Now, primarily, care-givers should be parents. Even foreign, many different men are used for making multiple babies. ... The end is near; it cannot be helped!”

3352, Mind your tractor

Apparently, someone during the previous messages in the conversation tree had complained that the young are driving tractors. In the message 3352 that complaint is mocked to be from someone with “boring retirement days.”

6364, Family Lahtinen
It was interesting that this conversation discussed a very atypical pensioner type, that of a pensioner with children. Otherwise the discussion had not much to offer for this thesis, although the last writer told sarcastically that they were happy to have something to spend their time with, as they were fighting with bureaucracy.

7584, Customer service
In the conversation, the competence of the older workers as customer servants is being disputed.

7875, Retired mason
The one message in the discussion told, the writer has used the services of a retired mason for reconstruction work.

8006, Hanging on with pain until retirement
The conversation went mainly on discussing the care for the elderly, so, although a topic of the elderly, it did not directly relate to retirement. Why it ended up in the data was, it mentioned an older nurse who was working in the elderly care. They were told to be hanging on in the job, waiting for retirement age while suffering pains.

8153, 28 and retired
What became the most cited discussion in the analysis was this monologue of a 28 years-old who was forced to be already on disability retirement. They had been ill for 7 years and wanted to go on in her/his life. The problem was, they could not find a profession that would have suited her/his fragile mind. The reasons they gave for wanting to end what other people were looking forward to, namely, freedom from work, were the too small pension they were receiving as well as intrinsic desire to be useful for the society.